


```

Db      1923 SEDDEEY -PDBVVAAPAPAPAFTEEBEVENERIROEBELKEVLQ-----QIMARGEVA 1972
Qy      2102 HLPHLRPLPESOPSSPFLQAPGVKQHOR---VVTIAOHISEV-----ITODY 2147
Db      1973 NDEPDVQMAKTTTSQTSTSTKPYVTAPKSDPIPIAPSOBSKIEBERIRTEALBEEEF 2032
Qy      2148 TRH-HPQOLSAPLPAPLXPSPGASCVLDLRRPSPDLYLRPPDHGAPARQSPHSSEGGKRS 2206
Db      2033 YRHGHNFLEBSP-----EEDEVSINMEDEVEYAEIARLYESANQTMRR 2074
Qy      2207 PEKPKTSVLGGEGEDGIEPVPBEGMTBGRHSRAV-----YLLYLRADGOTE 2253
Db      2075 PGPYTYTTEDESEB-----DGTLSNSESRLVAREKRLMDKKTADSLMAKQOMKKVQ 2126
Qy      2254 PSRWGSKSPGNTSOPPAF---FESKL-----TESNAMYKSKKOEINKQLANTHNREPE 2303
Db      2127 AKQTTTASSTSVTATPAVALINFSDLKTSRTTDSNVPFETTK-----NIPa 2173
Qy      2304 YNISQPGTEIENMALTGTGLMYTRS-QAYOEHASTNGLEAIRKLMGYDMOWESSPP 2362
Db      2174 LEIDDPKKDI---PEISASIDKTMAEVDALGOVYTN---EKALPNLCLFDQSNFSNHP 2228
Qy      2363 LSAVAFNPL-----NASASILPAAMPITAA-----DGRSDHTLTPSGGKG 2402
Db      2229 TSSASTSTADDLILKKNXNSSSPFLPLPGOSVLGSOQLDSVRDNRKXNENMTTSPG--- 2285
Qy      2403 AKVSGRPSSRRKAKSPAPGLASGDRPPSVSVSHSEBGDCNRRTPLTNRWEDRPSSAGSTPF 2462
Db      2286 -----LKRSPGMILPS-----PTSTSIFFMPMPTAAESVGA 2315
Qy      2463 PYNPLINRLQAGWMASSPPP 2482
Db      2316 AIGATTASMGGISVADPPP 2335

```

[illegible]

```

OY 286 RRRNRRKQKQKFCQRIDOLMEALKEKVEE1ENNRPRRAESKREY---YKQFPPELRK 342
Db 370 TINGQEOEOEOEOOQOQOQOQVKEEREDEI1EODDN-----LEVSDSDOKLEE--H 418
OY 343 QRELOEMQSVGQSGSLGSMARSHEVSEI1DGLSEONLEKQROLAVIPMTYDA 402
Db 419 QKELLEQONQ--NOEGS-----TDENEYTE---QEBEBEBDEEEOVISTPA---- 460
OY 403 DOQRKEITNMGMLADPMKYK-----424
Db 461 -KKRISF-----QDPTHTYEAVRTDDEDDDEDEEGDEYDHGDDEDEDEDE 513
OY 425 -----DROV-----NMMSQOEKETFREKFMQHPKQGLIASFLERKTVAE 465
Db 514 YDEDEDEDEDEKQIOYQOOLQYQOOLQKQOEQOEYQOQOOLQ-----554
OY 466 CVLYYYLTTKKNKYNKSLVRSYRRRGKSQOQOQOQOQOQOQOQOQPMPRSSQOEKEDEK 525
Db 555 -----LOREBEY-----QROLQOQOQOQOQOQOQOQOQHQQOQYDDDDDD 594
OY 526 EKEAEKEKEKEVENDKEDLLKEKTD--DTSGEDNKEAVASGKR--TANSQGRK 579
Db 595 DEEBEYDDVIRHDTSEBSKDXTP1P1MDHFKQKSEKVKAEQOEFNNVAVANSQETDQ 654
OY 580 GRITRSNANEANSSEAL-----TQOQSELASMELNBSRKTBEEMETAKGGLENG 631
Db 655 QOQOQOQPKHEEBEEPKVEDVKVETEBQTEKEEDVKVDESTVKDEVPDKT----- 708
OY 632 RNMGSAIARWGSKTVNSOONFYFNYKKRONDEL1IOOHKLMEKERNARRKKKAPAAAS 691
Db 709 -----VEESVEQV---AEPNKKVBEVBEKVAE-EERTVBEVKAEBPV--- 749
OY 692 EEAFFPVVBEDEMEASGVSGNEBEMVEEA---BALHASGNEVBRGCSGEPATVNNSSDT 748
Db 750 EEVKVEBPVEEVKAEBPVBEVKAEBPVEVKTETPVEVKVEBPVEEVKVEBPVEVEAE 809
OY 749 ESISPSRTEAKLDQNGQNGPKPALLGDGPPGPGPTPPRRISRARIPTTASANTGATP 808
Db 810 ESVOBEPVEEVKD-----BTKVBEPTKVEEP 836
OY 809 PPAPPSBPAPPVVPKEKEEBETAAPRVBEGEOQKPPAAELAVDTGKAEBPVSECTE 868
Db 837 -----1BEVKVEEPTKVES1BEVKVEEPTKVEE-PVEELKPEEPTKVE--E 880
OY 869 EAEBGPAGKDAEAATAEGALKAEEKGSGRATTAKSSGAPGSDSDSATCSADEYDE 928
Db 881 SVED--VKVEDVKVBEVKAEBEPTKAE-----SVEDVKVEBPIKVEBPVK 923
OY 929 AEGGDKRLLSPRSLTLTPTGDPPRANASPOKPLDKOLKQAAAA1PPIQV--TKNHPP 985
Db 924 VEEBPV-----VEBPVKVEBPIKVEBPIKVEBPIKVEBPIKVEBPV 970
OY 986 REDAPFTKAPAPPPONTLOPE-----SPAPQOQSGSPGKSPAPAPADKEAFA 1036
Db 971 KVEVA--SPVVQEOBPQOEKREPEVNVSTTITIASPOQSSNP-----BETPV----- 1016
OY 1037 AEAQKLPGBPDCMWTSGLEFPVPREBVIKASHPADPSAFSYPAGHPLPLGLHDTAREVL 1096
Db 1017 ----KQPOQGE1VNVSTFLK-----QOQOQOQGTPTQGTQTPFKQNHTEINWQTPTKTP 1065
OY 1097 PRPPTISNPPPLIS 1110
Db 1066 PLPQPTNEKTIPLIS 1079

```

RESULT 99
 S27923
 gene Lf3 protein - human herpesvirus 4
 CSpecies: human herpesvirus 4, Epstein-Barr virus
 CDate: 19-Mar-1997 #sequence, Epstein 19-Mar-1997 #text_change 31-Jan-2000
 CAccession: S27923
 R.Parker, B.D.; Bankier, A.; Satchwell, S.; Barrill, B.; Farrell, P.J.

```

214 VSPPEIESKRSIVQIITDENRKKAAEARILEGLQPVLELNNQPSDT-----RQY 266
      |||||:
206 V|PPPIASHPT-----PPSHKTDITGSRHSSGSA 236
      :
267 HENIKIQAMRKULILEFKSRNARKQWKFCQRYDQLMLEKKVERIENPPRRAKE 326
      :|||:
237 HSQGFSTPSISGKIFPDQAKTRTHLYVEI---KDEVIIVEDK--DKTEFKPPPNVS- 290
      :|||:
327 SKVREYKQFPEIKOREL-----QEMOSRVGORGSLMSAASEHVS 373
      :|||:
291 ---NELTAEPEHRIIRINEMAGIDEPIOPPIAOERKSSV---VSLGNMFVGVDHES 344
      :|||:
374 EIDIGLSEQNELEKQMOGLAVIPMLYDADQOIKFTLNNGMLADPMKYKQROYVNNMS 433
      :|||:
345 E---LTTSE--KSHIRMS-----LMAEKN---DEIILD--- 370
      :|||:
434 EOEKETEREKEMQHPKN-FGLIASFLERKTVABCVLYYLTKNENYKSLVRSTYRRCK 492
      :|||:
371 EQSK-----PKSTPGLKNPF-----GK 387
      :
493 SQOQOQOQOQO---QOQOQOQOQOQPPRSSQREKQEKKEKEKEKEKPEVENDKELLKE 548
      :|||:
388 ATQSVMAHTDQSVIKVNOQOSKQSLGULQEBLNDIAQTESAQOSKQELTQDEIDHT-A 446
      :|||:
549 KTDQTSGEDNDEKEAVASKGRKRTANSQGRKGRITRSMAENSEAATP---QOSAEI 604
      :|||:
447 RINMAADPFQOPAVNNMQGLSQEKD--HIARIEAMADDSOSKIIVPPPSIIQDIEL 504
      :|||:
605 ASMEINNESSR--WTEEMETAKKGL-----EHNMMALARMQSKTV---SCKXNF 652
      :|||:
505 SWEEMDHARIAAMADEDFTHPVKGAVAPIYDENYSRBDGAIIDFPASTATVPVFAQPSEI 564
      :|||:
653 YNNYKKQGNLEI-----LQGHKLM--EKERNARRKKKAPAAASE 692
      :|||:
565 ELSEEREHEISIAAMADEPNTVPVSTHPIQOIEITELTEKKHIAI---IEMASE 619
      :|||:
693 EAAFP-PVDEBEMEAS-GVSGN--EEMVEEALHASGENVPRCEGCPATVNNSSD 747
      :|||:
620 DINAASPFIQORAVATVPVPLNYHVEBPTLSOEIDHIA-----RITAMATED-TD 671
      :|||:
748 TESISPHTEAKQDQNGQPKRPATLGADGPPGPPTRPRRTSRAPLEPTPASEATGAPT 807
      :|||:
672 LGTLPPVOVK---QNEPE----- 686
      :
808 PPPAPPPSAPPVYVPEKEKEETAAPVVEGEQKPPAAEELAVDTGAEEVYKSECT 867
      :|||:
687 -----LSOEIDHITRIAMANEDFGMYPSIVSE-----HPAVESELT 725
      :|||:
868 EBAEBEPATKQAEAAEAATAGBALAKKEGSGGATKASGAPQODSDSATCSAEVD 927
      :|||:
726 QOEBLHIAK-----ITGMSNDITLPP----- 748
      :|||:
928 EAEGGCKNRLSPRPSILTPGDPPRANASPOKPLD-LKOLKORAAAIPIQVTKVHEPP- 985
      :|||:
749 ---PTGKPSBTSJLQEBLHIRIARIEMASAEYDVP-KIFEPPE 788
      :|||:
986 -----REBAATKPPAPPP----- 1001
      :|||:
789 LTOEBLHIAKITAMAAQDVOLPATQRSITTKHVSILTPPPPPSKHFEODLTQOEBLHIAI 848
      :|||:
1002 -----PONTQPSADAPQOQPSSPRGKSRPAPPADKEAPA-----EAQ 1040
      :|||:
849 IAEAMDMQNTPTPTAIPVQDEBEERPTTESGSATGADIPQOEBDASGASGSPDNNAQ 908
      :|||:
1041 KLQGDPPCWTSGLPFPVPVPREVITKASPAAPDSATSYPAPGHPRLPLGHDPARPLP--R 1098
      :|||:
909 VL-----TSG-----FSPDRVTSPPAALDTTEEPQPIWA 937
      :|||:
1099 PPRTISNPPRLISSAKGPNV-----LERQIGALISQG----- 1128
      :|||:
938 QKTVSPSPASDNASKRSSEYDIRSISEIROSESIDIGKMYTEQOJLSPFKQSIHDEEDVG 997

```

Qy	1129	-----	MSVOUHV	-----	PYEBHAKPVGPT	-----	MGJPLRMDPKLAP	1163																																						
Db	998	HEIRTDVEEFL	EVYEBOLHFL	RGIDVESQHNH	-----	PTTSAFRTGSR	ICEGEXKKG	1053																																						
Qy	1164	FSYKQOQL	-----	SPRGAGPRESL	-----	GVPTKQEA	SV-LRGTAIGSVGCSITK	1211																																						
Db	1054	DDAEQOQLENY	EEETKSSSR	READDDGE	FTQRESLAQQ	PTDLSL	PGSMYLRPNMG	1113																																						
Qy	1212	IP	-----	TRVPSDAIT	RGSIITHG	TPADVLK	GTITRLIGEDSPRLDR	1257																																						
Db	1114	FLSNIMADAI	INKAEAGSQIOA	AVPIPRPSSSSNN	IVNNVFFSSSKS	STSLGTSAPLK	---	1170																																						
Qy	1258	GREBSLPKHV	-----	IYEGKHGVL	SYEGMGSV	YQCSKEDSRSSGPHHETAPKTY		1311																																						
Db	1171	-----	SIPBQIGIPMD	GSIEBERQI	SVMAAD	FPDDSVNNV	PTSSSSNIPAGMEDLS	1226																																						
Qy	1312	DMMEGRVRA	ISSASIEGLMG	RAIPRRHS	PHNLKQ	HNIRGSI	ITQIGIPRSYEAQEDY	1370																																						
Db	1227	EAERKIMSV	ANAME	-----	MGARPP	-----	SSQIPTRSPVSSSI	IMELPRGLDLEDDER	1283																																					
Qy	1371	-----	LRBAKLKRE	GTTPPPPSR	DLTEYKQAL	GPLKLP	KAHGLVAIVK	1419																																						
Db	1284	NKIMAVMA	ADMOVNRK	IA	-----	GRPPMPSTN	-----	IPRGMEGL	---SEE	1326																																				
Qy	1420	EAGRSIHEI	PREBLH	TEPLPLAR	PKESIT	QGPRLKYDTG	ASTTGS	KKQDVRSLLGS	1479																																					
Db	1327	EQOKIMSV	MAA	ETDSSSV	ITSTQPS	RSPSVAMQ	-----	QLMP	1367																																					
Qy	1480	PGRTPEPVH	-----	LDVADARAL	ERACEYES	LKSRPCTAS	SSGSIARGA	1526																																						
Db	1368	POQAIP	IPRPLEGL	SDERHT	IMSVAAEA	-----	FEESRSOV	PSRQPSRS	1414																																					
Qy	1527	PVIVELK	PROSP	PLYTDH	GAP	-----	FAGHLPRGS	PVTUREPT	PLQEGSLSS	SKASQDRKL	1585																																			
Db	1415	-----	PSFVNPQS	-----	FHPIRPE	PIVPRGL	EDLS	EEBEOQKIMSV	MAAEV	EEBSRQL	1465																																			
Qy	1586	TSTPERIA	KSPHSTV	BEHPRH	PISEYH	NLG	VCV	-----	DLYSHN	PLAFDPT	IPRGI	1641																																		
Db	1466	PS	-----	RQPSRSP	SVAMIO	APAVIIP	-----	SGMEDL	PAERQOKIMSV	MAEA	ETQNV	1515																																		
Qy	1642	PIDDAAYL	PRHLANP	TYPLH	VPYLRG	VPYAL	ENRQ	ITINDY	ITSQ	-----		1693																																		
Db	1516	PSRSPSNY	-----	SMQDV	VIYTH	-----	GLEDS	SEAE	ROKILSV	MAEA	ETD	AKIPS	1562																																	
Qy	1694	-----	OMHNAT	ATAMA	QADML	RGLSP	RESSL	ALNTA	AGRG	ITDLSQV		1737																																		
Db	1563	RSTSSYS	MPRLPQMS	QPEIT	TGLEH	SEAD	MEFG	-----	RDBS	-----	RSHQV	IPGLE	NS	1617																																
Qy	1738	PHLPV	VPPTPG	TPATMD	RLAYL	FTA	-----	POFSS	RHSS	PLSPQ	-----		1781																																	
Db	1618	EROQIM	-----	SVMAHA	LESIT	PSGH	IDPISL	PRGH	TGPK	AGI	VNEDEL	PETERKQ	1671																																	
Qy	1782	-----	PTHLKPT	UTSSERE	-----	RD	-----	RDER	RD	-----	RRRE	1811																																		
Db	1672	RESPT	REGVAT	ISYER	ELAM	QBER	MDGL	IEDI	IRIR	EGAS	RSD	RSDE	VLR	REED	1731																															
Qy	1812	KSILT	-----	STTVEH	APIM	RPCTE	QSSSSSSG	GGSSSR	ASHNA	HOHSP	IS	IPRT	1867																																	
Db	1732	PEVHTE	BEEST	AVTVDV	PSVSPVTE	VEPER	QOTDD	FTYSS	RFE	LI	YQO	EEB	AGSLQ	1791																																
Qy	1868	QDLQOR	PSVLA	HTGK	GIIT	AVES	KFVL	STST	SSPVP	ATF	PTAT	CP	LG	ITDG	1937																															
Db	1792	KQKDE	EFKME	-----	IVPFG	DSEL	PHQ	OFV	NE	PTTK	-----	KTSDF	1831																																	
Qy	1928	VUTPL	MEPV	-----	LPKAP	RVAR	REPR	ADTG	HAFL	KPR	AGSL	EPAS	PSK	SEPR	PLVP	1966																														
Db	1832	DFP	ET	DEV	EK	SE	ELOR	IVTK	MD	VDD	DE	Y	-----	DNV	IA	TEP	ASSV	QSR	---Q	1881																										
Qy	1987	PVSG	HA	-----	TA	RP	AKUL	AP	HN	AS	RO	PAR	PA	S	AD	PH	NE	K	Q	SP	FS	IO	EL	ET	LSL	2041																				
Db	1882	PV	SET	SV	KRT	IQ	RG	PTK	-----	PP	MI	K	T	Y	E	EL	TKS	-----	DE	DE	SSC	1922																								
Qy	2042	G	YH	SS	S	P	G	V	P	V	S	P	S	L	T	D	K	L	P	K	H	L	E	I	D	K	S	H	L	E	G	E	L	R	P	K	O	P	G	V	K	L	G	E	A	2101


```

QY 1011 APQGGSSPRGKSSPAPADK-----EAPAAEAQKLPEDPPTCTSGLPFVPPREVITKA 1065
Db 369 QPBGPGLPVPGQAGAPGPPGGERGKDRGFPGTS--LPG--PSGRDGLPG-----415
QY 1066 SPNLPDPSAFVAPGPHLPGLHDTAPVLPRLPTISNPPPL-----ISSAKH 1114
Db 416 -PGSD-----GPPGDP--GYTNGIVECORPPGGDGGPFGIIPQOPFGIEIGKGGK 464
QY 1115 PSVLEROI GAI-----SQGMSVQLHVPYSEHAKA-----PVGPTVGLPLPMDPKLADPS 1165
Db 465 ESCLICIDYGRGPPGPGPCEIGFPQOPGAKGDRGLPGRDVGAVGPGQCPGLIQP 524
QY 1166 GVXGE-----QLSPRGAAGPPEISGVPTAOEASVLRGTALGVSQGSITIGISTRPS 1219
Db 525 GAKGEPGFYDLRLKDKGDPGPPGDPG-----MPRA--GS--PGRDGHPLPGPK--GS 575
QY 1220 DSATYRGSITHTGPAVLYKGTITRIIGEDSPSLDRGREDSLPKGHVYE-----GKK 1274
Db 576 PGSVGLKGB--RGPPGVGFPGS-----KQDTGPPGPPGPPGAPIDDK 617
QY 1275 GHVLSYEGMSVTQCSKEDGSSSGPPHETA PKRTYDMEGRVGSAISIEGLMG-- 1332
Db 618 G-----QAGFPGGPSPGLPGPKGP-----GKIYVLPGPAGBGLPGSP 657
QY 1333 -----RAIPREHNSPHLKEOHNGISTQ--GIPTSYVEAOBYLREACTLRE 1381
Db 658 GPPGQDGRGFPFGTGRPLPGE-----KGAVGQPGI GPPGPPGVGLPDD--MPP 710
QY 1382 GTPPPP--PSRDLTAAYKTQ--ALGPLKLAHGLVATVEAGR--SIEHIREE- 1432
Db 711 GTPGRFNGNLPNGNNGVGGKGPBGVGLPGLKGLPGLPGIIPPTPEKKSIGVPGVGEHG 770
QY 1433 -----LRHTPELPLAPRLKEGSI--TGSTPLKYDTGASTGSK--KHDVSLIG 1478
Db 771 AIGPPGLGIRGEPGPPGLP-----GSVGSPPGV--GIQPGARPGGQGPGLISG 820
QY 1479 SPG---RTFPPVHLDVMA-----DAPALERACTEESLKSPPGTAASSGGSIMAGAPVI 1529
Db 821 PPGIKGEGFPFGPLDMPGPKDKGAQGLPGITGSGLPGLPGQGAAGIIPGPPSGKE 880
QY 1530 VPELCKPPOS-----PLTYEDHAPAGHLPGSPVPTME-----PTP 1567
Db 881 MGWMTPPQPGSPGVGAPGLPGEKGDHGFPGSSG--PKQDPLKDKDVGLPKPGSMD 939
QY 1568 RLQEGSLSSSKASQDK-----LTSTPREIAK-----SPHSVPEHHPH--- 1606
Db 940 KVDMSMGQKQDQEGKQIGPIGEGKSAGDPGTGVPCKDQAGQPGQPGKDPGISG 999
QY 1607 -----PISPYELLRGVSGVDLYRSHIRPLADPTISIPGILPDA 1647
Db 1000 TPGABGLPQPKGSVGKGLPGTPG--KGVPGIP-----GPGSP--GLFGDKGA 1045
QY 1648 AYVLPRLHAPNPTTYPLYPVLYRGVPTALENQTIINDYITSQOMHNTATAMAORA 1707
Db 1046 KGEKQAGPFGIGIGL-----RG-----EKG 1067
QY 1708 DM-LRGL--SPRESSLALNYAAGPPIIDLSQVPLPVL-----VPPGTATAMD 1756
Db 1068 DQIGAGPFGSPGK-----GEGKSIGIPGMPGSPGLKSGPVGVPGLGEGEKD 1119
QY 1757 RLAYLPTAPQ--PSSRHSSPLSPGAPTHLTCTTSSSERRDRDRDREREKSL 1815
Db 1120 K--GLPGLDGI PGVKGAGLPETPG-----PTBPAGQKGPESDGI PGSAAGEGPEPL 1170
QY 1816 TSTTVEHAPVWRPCTEOSSGSSGSSG-----GGGSSSRPASHSHAHSHSISPTODA 1870
Db 1171 PGRG-----PFPKAKGDKSGKGVGFPGLGSPPIIPSGKBOGMGPPGQGP 1222
QY 1871 LQGRPSVLHNT-----GMKGIITAVEBSKPTVLASTSTSSPVNPATTPPATCPLG 1923
Db 1223 LPGSPG--HATGPKGDRGPGQ-----QPGLPGL-----PGWGPFG-----LPG 1260

```

```

QY 1924 TLQGVPTLMEPLLVPKBAPEVARPERBRADTGAFLAKPPARSGLEPPSSKSEBRP 1983
Db 1261 -IDGVKGDKNP-----GMPAGVPGPKGDPG--FGCMPI--GSGGIGTSGKDMPP 1310
QY 1984 LVPPVSGHATARTP-----AKNLAPHNASDPDAPAPASADPREKTOSKP 2030
Db 1311 GVPGFQPKGLPGLQGIKGDQDQVPGAKGL-----PGRPPG-----P 1351
QY 2031 FSIQELRLSLGYHGSSTPSPGVPEVPSPSLTHDKGLPKHLELDKSLBEGELRK- 2089
Db 1352 YDIIRGE--PGLPG--PBG-----PBGKGLQGLP-----GPKG 1381
QY 2090 QGFPVKGGEAAHLPHLRPLESOPSSPLLQTPAGVGHQVVTLAQHISEVITQDYTR 2149
Db 1382 QQGVTLGVLG-----IF--GPPGIPGDPGAPQKG-----1408
QY 2150 HHPOOLSAPLPAPLYSPGASCPLVILRRPPSDLYPPPDHAPAGSPHSEGGKRSPEP 2209
Db 1409 ---EMGPAGPTGRRGPGP-----PQPD-----GLPGSMGPPGTPSV 1442
QY 2210 NKTSLVGGGEGDIEPVSPPEGMPERHSRAVYPLVYRDG--EOTBPSMGSKSP--GNTS 2266
Db 1443 DHGFLVTHSQITDDPQCPGSGTKILYHG--YALLVQGNERRAHGQDLGTAGSCLRRFS 1498
QY 2267 QPAPFSLTBSNSAMVYSKKOEINKLNTNRNREPEYNISQPTGFI PNMPTATGTGLMT 2326
Db 1499 TMPPLFC-----NINNVNCPASKNDYSIWLSTBERMPSMAPITGENTRP 1543
QY 2327 YRSQAVDHASTNNGLEAIIRKALMKYDQWESPPISANAFNPLNASLSLPAAMPITAA 2386
Db 1544 FISRCVACEADA-----WMAVHVSQTIQLP--CPGSMWSIMIGSYFM-----1585
QY 2387 DGRSDHTLTPGGGGKAKVSGRPSRKAASAPGLASD-RPPSVSVSHSGDCN 2440
Db 1586 -----HTSAGAGSGQALAS-----PSCLEFRSAPFIECHGRGTGN 1623

RESULT 97
725752
Hypothetical protein F45B4.4 - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C.Accession: T25752
R.Wilson, R.
submitted to the EMBL Data Library, September 1996
A.Description: The sequence of C. elegans cosmid F45B4.
A.Reference number: 220082
A.Accession: T25752
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-2361 <MIL>
A.Cross-references: EMBL:U70852; PTDN:AA09135.1; GSPDB:GN00022; CESP:F45B4.4
A.Experimental source: strain Bristol N2; clone F45B4
C.Genetics:
A.Gene: CESP:F45B4.4
A.Map position: 4
A.Intons: 60/2; 111/2; 939/3; 977/2; 998/1; 1021/1; 1125/3; 1140/3; 1157/2; 1173/2; 131.

Query Match 2.4%; Score 319; DB 2; Length 2361;
Best Local Similarity 17.6%; Pred. No. 0.00068;
Matches 500; Conservative 355; Mismatches 1017; Indels 968; Gaps 118;

```

A: Molecule type: protein
 A: Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GIA>
 A: Experimental source: placenta
 A: Note: the amino end of the mature form is blocked
 R: Soliman, R.; Haka-Ristku, T.; Prockop, D.J.; Trygsvaen, K.
 FEBS Lett. 225, 188-194, 1987
 A: Title: Complete primary structure of the alpha(1)-chain of human basement membrane (type IV) collagen
 A: Reference number: S00207; MUID:88083584; PMID:3691802
 A: Accession: S00207
 A: Molecule type: mRNA
 A: Residues: 244-530 <S013>
 A: Cross-references: EMBL:Y00706, NID:G29546, PIDD:CAA68698.1; PID:G29549
 R: Eble, J.A.; Golbl, R.; Mann, K.; Kuenn, K.
 EMBO J. 12, 4795-4802, 1993
 A: Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
 A: Reference number: S39614; MUID:94038963; PMID:8223488
 A: Accession: S39614
 A: Molecule type: protein
 A: Residues: 371-554 <EBL>
 R: Babel, W.; Glanville, R.W.
 Eur. J. Biochem. 143, 545-556, 1984
 A: Title: Structure of human basement-membrane (type IV) collagen. Complete amino-acid sequence
 A: Reference number: A02863; MUID:85003629; PMID:6434307
 A: Accession: A02863
 A: Molecule type: protein
 A: Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-
 A: Experimental source: placenta
 R: Glanville, R.W.; Rauter, A.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
 A: Title: Peptide fragments of human placental basement-membrane collagens showing interrupted
 A: Reference number: S16908; MUID:82005835; PMID:6792033
 A: Accession: A58517
 A: Molecule type: protein
 A: Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-14
 R: Wachtel, R.S.; Benson, V.A.; Lovell, K.T.; van der Rest, M.; Fietzek, P.P.
 Biochemistry 22, 4940-4948, 1983
 A: Title: Isolation and characterization of pepsin-solubilized human basement membrane (type
 A: Reference number: S16910; MUID:84053346; PMID:6416291
 A: Accession: S16910
 A: Molecule type: protein
 A: Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-
 A: Experimental source: placenta
 R: Philanthem, T.; Trygsvaen, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; F
 J. Biol. Chem. 260, 7681-7687, 1985
 A: Title: CDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
 A: Reference number: S01466; MUID:85207819; PMID:2581969
 A: Accession: S01466
 A: Molecule type: mRNA
 A: Residues: 1256-1669 <PIH>
 A: Cross-references: EMBL:M10940, NID:G180421, PIDD:AAA52006.1; PID:G180424
 R: Brinker, J.M.; Gudas, L.O.; Loidl, H.R.; Wang, S.Y.; Rosenblum, J.; Ketallides, N.A.;
 Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
 A: Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
 A: Reference number: S16879; MUID:85216555; PMID:2582422
 A: Accession: S16879
 A: Molecule type: mRNA
 A: Residues: 1259-1663 <BRI>
 A: Cross-references: EMBL:M11315, NID:G180817, PIDD:AAA52042.1; PID:G180818
 R: Oberhauser, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
 Eur. J. Biochem. 147, 217-224, 1985
 A: Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A: Reference number: A02864; MUID:85127033; PMID:2578961
 A: Accession: S19051
 A: Molecule type: protein
 A: Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',
 R: Siebold, B.; Deutzmann, R.; Kuenn, K.
 Eur. J. Biochem. 176, 617-624, 1988
 A: Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
 A: Reference number: S02550, MUID:89005112; PMID:2844531
 A: Accession: S02550
 A: Contents: annotation; disulfide bonds
 C: Genetics:
 A: Gene: GDB:COL4A1
 A: Cross-references: GDB:119791, OMIM:120130

A: Map position: 13q34-13q34
 A: Intron: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/
 1; 731/3; 782/1; 820/1; 876/1; 967/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 116
 C: Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2
 C: Cations among trimer amino-terminal domains (disulfide and desmosine cross-links), di
 C: trimer associations in the interrupted helical domain (with disulfide and desmosine cr
 C: Function:
 A: Description: structural component of extracellular basement membrane
 C: Superfamily: collagen alpha 1(IV) chain
 C: Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplicatio
 F: 1-26/Domain: signal sequence; #status predicted <SIG>
 F: 27-1669/Product: collagen alpha 1(IV) chain; #status predicted <MAT>
 F: 29-1669/Domain: amino-terminal nonhelical, 7S <7SD>
 F: 163-1440/Domain: interrupted helical <COL>
 F: 414-452/Region: integrin binding; #status experimental
 F: 597-599/Region: cell attachment (R-G-D) motif
 F: 917-919/Region: cell attachment (R-G-D) motif
 F: 968-970/Region: cell attachment (R-G-D) motif
 F: 1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F: 1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTR>
 F: 1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>
 F: 172-540/947/Modified site: 5-hydroxylysine (lys) #status atypical
 F: 172-645/839/Modified site: 5-hydroxylysine (lys) #status atypical
 F: 446-447/Cleavage site: Gly-Ile (gelatinase B) #status predicted
 F: 766-775, 784, 787, 790, 796, 799, 804, 810, 816, 822, 834, 860, 863, 869, 872, 875, 887, 890, 893, 899, 90
 23, 1169, 1138, 1141, 1159, 1171, 1176, 1179, 1194, 1200, 1215, 1224, 1227, 1244, 1247, 1250, 1256,
 431, 1437/Modified site: 4-hydroxyproline (Pro) #status experimental
 F: 1120, 1268/Modified site: 5-hydroxylysine (lys) (partial) #status experimental
 F: 1120, 1268/Binding site: carboxylate (lys) (covalent) (partial) #status experimental
 F: 1212, 1424/Modified site: 3-hydroxyproline (Pro) #status absent
 F: 1392, 1395, 1398, 1404/Modified site: 4-hydroxyproline (Pro) #status experimental
 F: 1466-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F: 1505-1511, 1616-1622/Disulfide bonds: #status predicted
 F: 1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 2.4%; Score 320.5; DB 1; Length 1669;
 Best Similarity 21.0%; Pred. No. 0.00041;
 Matches 415; Conservative 133; Mismatches 750; Indels 677; Gaps 100;

QY	728	GNFVREGCSGPATVNNSSDTEISPSPTTEAK-DTQNG-----PKPATGADGPPPG	782
DB	64	GNQGBRG- GGR- -GQKCDTGPPLPTKTRGPGASGVGNGLGIRGQD-PPGP	119
QY	783	PTPP- -RTSRPIETPPVASEATGATP	808
DB	120	PGIPGNGTGERGPGLPFGAGNPGPGLPGMKDPPGILGHVGMILKGERGPG	179
QY	809	-----PPAPSPAPPVV-----PKEKEETAAAP	835
DB	180	IGTPPGPPLPGIQGVPGPGTGPGRPGPPGPPGKGMKLSTQGRKGDGDGVSGP	239
QY	836	PVEGEGRKPPAAEELAVDTGAKEPVSECTEAEG- -PAKQDAEAATAGAL	891
DB	240	PGVPGAG- -VGE- -KGDFATKGEKQKGPFGQMPGVGEKGPGRG- -	287
QY	892	KAEKKEGSGRATTAKSSAPDSSATCSADVDABEGGNLLP- RSLTLPGD	950
DB	288	-----KPKGKGDGKEKSPFPGEPVGLGR- -QGQGEKGAAGPPGIVIGT- -	338
QY	951	PRANSPQKPLDLKQKQAAALPPIQVTKVHPREDAPKPPAPPPQNLQPSD	1010
DB	339	-----PLGK- -GERVPTGPGRGPGKGRPGLP	368

```

Db      726 EQAVEADAGPTGANE-----DDSDVPAVPLSEVDANVERMEQQAOKSSEQ 773
Qy      843 QKPPAEELAVDTGKAE-----EPVSEK- 866
Db      774 PEQAAATEVSKELSESQVHMMMAAVALDGTAAITIEERSPSMISASVTEPLEQVEAEAL 833
Qy      867 -TEE-----AEGCP-----AKGDAEAAEATAEALAEKKEGSGGATTAKSQA 911
Db      834 LEEVELEREVIASEEPPTVTEPLPENREARAGDVVSEALTEP-----AVTAATFAG 885
Qy      912 PQDSDSATSCADEVDAEGDKURLSPRSLPTPTGDPANASPOKPL-----DLKOL 966
Db      886 P-----LGAEBGTEASAAEBETTEMWASVQLTDSQDTTEATPVEVEGVDIDEQ 937
Qy      967 KQRAAALPPIQVTKVNEPRBDAAFTKPAAPPAPPPONLQPSDAPQPGSSPRKSRSP 1026
Db      938 ERRTQEVLAQVAEKVESQ-----LPGTGGEVDILQ-----PVQRAEAERPEQAE 984
Qy      1027 APPADKE---AFAEAQKLPQDPPCWTSGLPFPVPRREVIAKSPAPPSAFSTAPPGHP 1083
Db      985 ASGKKEEDVVLKVDQAEAKTEP--FTQG-----KVGQT-----TPESFEKAP 1026
Qy      1084 LPLGLHDARVLPRLPTISNPRLISSAKHPSVLERQIGALISQMSYQLHVPYSEHAKA 1143
Db      1027 -----QVTESESSE----- 1036
Qy      1144 PVGPVTMGLPLPMDPKKLAPSGVKQEQQLSPRGQAPPEISGLVPAQASVLRGTALGSV 1203
Db      1037 -----LVTTQCOETLA-----GVKSGQWME-QALIPDSVETPTSDSTDPVADDA 1084
Qy      1204 PGGSTTKIPISTRVPSDAITRRGSIHTGTPADVLYKGTITRIIGEDSPSLDRERESL 1263
Db      1085 PGTQKQKQEIIVIHENEVA---SGTQSGTEAEV---PAKKEP---DPA 1125
Qy      1264 PKGHVYEGKK-----GHVLSYEGMSVYQCKEDGSSSGPPHETAPKRTYDMM 1314
Db      1126 PSSFPQSETEKQSMEDTLEHTDEKVEVETVSIISKTEGQEA---DQVADETKXV- 1180
Qy      1315 EGRVGRALISASIEGLMGRAIPPERHSPHHLKEQHHRIGSTQGIPRSVEAQEDYLRE 1374
Db      1181 -----PPFEGC-----EGSIDTGTIVREKTEVALKGE 1209
Qy      1375 -----AKLLKREGTPPPPPSRDL---TEAYTQALGPKLKAHGLVATYKE 1420
Db      1210 GTEAECKDDALEIQSHAKSPSPVEREMVQVREKTEA-----EPTH----- 1254
Qy      1421 AGRSIHEIPRELRTPPLAPRLPKGSGITQGPLKXDTGASTTGSKGHDVRSLLGSP 1480
Db      1255 -----VNEKLEHETAVTVSEVSKOLLQTVNVP-----IDGAK--EVSSLGSP 1298
Qy      1481 GRTEPPVAPLDVMDARALERAQYSELSKSRPG---TASSGSGIARCAPVYBELKCP 1536
Db      1299 -----PPCLGQF-----EAVCTKIQVQSEASFTLTAABEERKVL--GETANILETGS- 1344
Qy      1537 ROSPLITYDHKAPRAGHLPKRGSPVTKREPTPLQSGSLSSSAQSDRLTSPRIASAP 1596
Db      1345 -----TLR-----PAGAHVLEB-----KSSBKND--FAALPGEDA--- 1374
Qy      1597 HSTVPEHHPHPISEYENHLRGVSGVDLYRSHIPLAFDPTISRGPLDAAAAYVLRHIA 1656
Db      1375 -----VF-----TGPDQAKSTPVIVSATV--KKGSSD----- 1401
Qy      1657 PNPTVPHLYPYRLNGYPDTALENRQTIINDYITSQOMHNTATAMAGRADMLGLSPR 1716
Db      1402 -----LEG-EKTSLKWKMSDEVQVACQEVKVALEIDLEPENGILELTK 1447
Qy      1717 ESSSLANYPAGRGIIIDLSQVHLEVLVPPPTGTAITMDRLAYL--PRAPOFSSRRHSS 1775
Db      1448 SSKLVQN-----IIQ-----TAVDQVTEETEAATELTLSELQV 1481
Qy      1776 PLSPGPHLTKPTTSSSERDRDREREREKSLTSTTVVHAHPIRPGTEQSS 1835

```

```

Db      1482 -----AHVYKADS-----QDAQGETKEKEEPLASQ---DETPTSKESES 1522
Qy      1836 GSSSGSGGGGSSSRPASHAHQHPISPTQDALQORPSVLNTHMGKGIITAVEPKP 1895
Db      1523 TAVQA-----HSDISKMSSEASE-----KMTVEVE----- 1549
Qy      1996 TVLRSTSSPVRAATFPATHCPLGTLGVVPTLMEPVLKPE-----APVAPRP 1950
Db      1550 -----GST-----VNDQLEEVVLSEEGGAGTKSVPE- 1579
Qy      1951 PRADTGHAFAPKAPRAGLEBPASSPSKGSBERPLVPVPSGHATIAFTPAKXLAHPHNASPD 2010
Db      1580 ---DDGHALLAEFLKSLVED--KEDEKGD---VDPEENQVSLADTDASG--GLTKESP 1631
Qy      2011 PPAPPASDPHREKTSQKPSSTOELRLSLGTHGSSVSPGCVPEVSVSSPSLTHDKL 2070
Db      1632 TNGP-----KQKEKEDN---QEVLEQEKVHSES-----DKAI 1661
Qy      2071 -PKHLEELDKSHLE 2083
Db      1662 TPQAOEHLQKQERE 1675

```

RESULT 96

```

CGHUB
collagen alpha 1(IV) chain precursor - human
MAlternate names: procollagen alpha 1(IV) chain
CSpecies: Homo sapiens (man)
CDate: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999
CAccession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58
J.Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A>Title: Structural organization of the gene for the alpha-1 chain of human type IV coll.
A.Reference number: S16876; MUID:89340433; PMID:2701944
A.Accession: S16876
A>Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-1669 <S011>
A.Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAA53098.1; PID:G180803
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1988
J.Soininen, R.; Huotari, M.; Hostlikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A>Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A.Reference number: A92680; MUID:89034231; PMID:3182844
A.Accession: A32117
A.Molecule type: DNA
A.Residues: 1-28 <S012>
A.Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233
R.Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A>Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c
A.Reference number: S02738; MUID:89030632; PMID:2846280
A.Accession: S02738
A>Status: translation not shown
A.Molecule type: DNA
A.Residues: 1-6,'L',8-28 <POB>
A.Cross-references: EMBL:X12784; NID:G30072
R.Birzel, D.; Oberhauser, I.; Dieinger, H.; Babel, W.; Gnanville, R.W.; Deutzmann, R.;
Eur. J. Biochem. 168, 529-536, 1987
A>Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem
A.Reference number: S00048; MUID:88029471; PMID:3311751
A.Accession: S00048
A.Molecule type: mRNA
A.Residues: 1-318,'A',320-944 <BRA1>
A.Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067
A.Accession: S25826
A.Molecule type: protein
A.Residues: 271-318,'A',320-554 <BRA2>
R.Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risceli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A>Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (75
A.Reference number: A23115; MUID:86004708; PMID:4043082
A.Accession: A23115

```



```
Db 3167 DLDHTPIA-----DHANVYHHGDSRIEKHASP-----VA 3196
Qy 1678 ALENROTIINDYITSQMHNTATAMAGRAMDLRGLSPRESSLALNVAAPRG-IIDLSQ 1736
Db 3197 TLEKKOKKVK-----SPKKEKVDPAPPEKKEIRLARVTEPEAEIIVDPV 3243
Qy 1737 VPH-----LPVL-----VPTPGTATMDRLAIVPTAQPF-----SRH 1772
Db 3244 VKETKEKKGDKSKLPFGFPFSKTKTKTKTGYPTEAAYTGEDLTDTDRHDLGCTSEH 3303
Qy 1773 SSSPLSPGPHLTLPKPTTSSSR-----RRDR-----R 1802
Db 3304 GEHPAFESSTISITTEADPDHEKVIKPSRPHMFSRMRHGEDDEVKPEGYILSSRY 3363
Qy 1803 ERDRDEREKSLTSTTVEHAPITWRPGTEQSGSGSGSGSGSSSR----- 1850
Db 3364 EGPDNNRPIGELLETPLDHSQVYH-----NGSGKLEKRSKKEKIKTPSPRVPSP 3417
Qy 1851 -----PASHSH-----AHQHSISRTQ-----DALQR-----PSVLAHTGM 1883
Db 3418 EIKLVARIQPLEHEDSEISKLSFVKDRSRAPSPSMFHROS KQGYEISPLYEGNLDV 3477
Qy 1884 KGIITAVEPSKPTVLRSTSTSPVPAATPPATHC----- 1920
Db 3478 TGRAAVE-NTPIEVRSVPEYHPKQEVSAQHPDESSQPRFRLIARMRHGEDDEIEKPD 3536
Qy 1921 -----LGGLDGVPTLMEPVLLPKFAPRVARPERPRAD-----TGAFLA-----K 1962
Db 3537 AYKFVTEVEYEGPLDIDRPV-----VELAQPISDHSGVYHNGSGWKKEKNIK 3583
Qy 1963 PPAASGLEPASPSPKSGSEPRPLVPVSGHATITARPAAKMLAHPNAPSPP-----PAP----- 2015
Db 3584 PP-----KPVNTKKGSLSPSEPISEVHEIRLITRVAASSSEBEVAVIIPASTVPKPERL 3638
Qy 2016 -----ASADPHREKTQSKP-----FSIOLELSLIGY-HGS-SYSP----- 2050
Db 3639 RFLSIGKKSPTSDESEFECSDQYDQVDTLSRSELEVMRYLHSGSVPPYSFKFTPVKVV 3698
Qy 2051 -----EGVEPV-----SPVSPSLT-----HDKG 2069
Db 3699 DKADERKKTIVYFRHGEEDFADKPEAYGFSDDVTGSLNEISRSELEHAIHDS 3758
Qy 2070 LPKH-----LEELKSHLEGELRKO----- 2090
Db 3759 QVYHDGLSYKIDNKKOKSKDRTELEBPTSRVAVLPOQESADSGSPKODKDGFRFR 3818
Qy 2091 -----PGPVKLGEAHLPHL-----RPLPESQPESSPLLOTAPGVK 2127
Db 3819 FFSKASPKSGPYEVPSPTEGPDIDVTGRVSELEHVPLOSREFPTTSPKKS--QEKPESEK 3875
Qy 2128 GHORVVTIAOH--ISEVITODYTRHPOOLSAFLPAPLYSPGAS-----CEVLDIR 2177
Db 3876 SSERVEVERKRYHLARI-----RHGEDEVENP-DTYGFASTYDGLBEFTSKSVDE 3928
Qy 2178 RPPSDLYLPDPDHGAPARGSPHSEGGKSPENPKTSVLGGGEGDGLIEPVSPP--GMTPEBG 2235
Db 3929 ETPISVYENVVHHGSGSIRVEKORAKISABEK-----DDDENVPKPSRFSPLRG 3981
Qy 2236 HSRSAVYVLYLRDGEQTEPSRWGSKSPGNTSQPAPAFSKLTES-----NSAMWKSKEQIN 2291
Db 3982 QKRVLSYSTLYSG-----PLETTRDLLENIPLSIVYHSGSDQHAHKRYLI 4029
Qy 2292 KKLNTNHNPEEYINSQPGTEIFNNPAITGTGLM--TYRQAQVQ-----HAATN 2339
Db 4030 ARIR-HAODEDEVAHKLHPVYSFPAYQSGILEKTLCKEVDVPIGRFARVYHSGS 4088
Qy 2340 MGLEIIRKALMGKIDQEBESPLSANAFLNANASLPAAMPITAAQGRSP----- 2391
Db 4089 GGLIWNLLKNRNSRASKMDEPKRRKQKQDKIKLILAVLPMKTEPKSDRFEKATKE 4148
Qy 2392 -----HTLTSGGGGAkakvgrpsrrkakspapglagadrrpvsyvhsedc 2439
```

```
Db 4149 EPLREAEVLRSTQSEAQNISTVSIPLSLTYDVP-----RPARIEGYTGDC 4197
RESULT 94
A45344
Immediate-early protein - suid herpesvirus 1 (strain Kaplan)
C:Species: suid herpesvirus 1
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C:Accession: A45344
R:Vilek, C.; Kozmik, Z.; Paces, V.; Schlim, S.; Schwyzer, M.
Virology 179, 365-377, 1990
A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented op-
A:Reference number: A45344; MUID:91021039; PMID:2171211
A:Accession: A45344
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-1446 <VLC>
A:Cross-references: GB:M4651; NID:G334070; PIDN:AAA47470.1; PID:G334071
C:Superfamily: herpesvirus immediate-early protein IE175
C:Keywords: DNA binding; early protein; transcription regulation
Query Match 2.4%; Score 323; DB 1; Length 1446;
Best Local Similarity 20.5%; Pred. No. 0.00029;
Matches 344; Conservative 158; Mismatches 590; Indels 582; Gaps 78;
Qy 686 AAPAASEAAPPVVEDEMEASGVSGNEEMVEAEALHAGNEVPRGEGSCGATVANS 745
Db 72 AAAAGATRRPPRPFAQQQQQPRRG-SGEIVLVDDEDEDEDEPGS--PAAGSPGRALHGS 128
Qy 746 SDTESIPSPHTEAAMDTONGQPKPP-----ATLGADGP-----PGP 782
Db 129 EHGHLVIGPRSRAG-----SGPRPPTPAAALAAAGAFGCGRSSPAAASPASSGSPGP 183
Qy 783 PTPPRRTSRA--DI-EPTPASE-ATGAPTPPA-----PPSPAPPPVVP- 823
Db 184 SAAPRRMGPARGDVPGBPGAPARPTPAPPAVAAPARBPASPSAPAGVSAFG 243
Qy 824 -----KEKEEETAAA-----PVYEGEGQKPPAAEELAVD 854
Db 244 GGAAPSAGDGRHHQHREPLDPPAARLLDRPLGARSVPSNPNNSNSTTTVAE 303
Qy 855 T-----GKAEE-----PVKSECTEAEEG-----PAKGDAEAAEATA 887
Db 304 TVAKCPEDDGLGLAGGAPPGQRPQRRAAGALRGKGFSSSSGSDSLSPARS 363
Qy 888 EGALKAERKEGSGGATTAASSGAPQDSATSCADEVDEAGEGDKNRLISPRSLTP 947
Db 364 PSAPRA--PAAARRSASSSSSSSSSSSSSSSSSSSSSGEEDEG-----VRP----- 407
Qy 948 TGDPRAANSPOKPLDKLQKORAAAIPIQVTKYHPRREDAPRTKAPRAPPPQNLQ 1007
Db 408 GAPIARAGP-----PSPPAPAAAPRASASA 434
Qy 1008 ESDAPQCGSSPRGKRSRA--PPADKEAFPAEAQCLPDPQMTSGLFPVPREVIRAS 1066
Db 435 SSSAASAPAP-----EPAPPRKRKSTNNHLSIMADGPPIDG--PL----- 477
Qy 1067 PHADPASFAYAPGHPPLGLIHTARVLPRLPTISNPPLISAKAPSVLERQIGAI 1126
Db 478 -----LTLGEPMP-----GSDPPADGVRVYGAAGDSREGIMD 510
Qy 1127 QGMSVQLHVPVSEHAKAVGVYTMGLPLPMPKTLAP-----PBGVQEQOLS-----PRG 1176
Db 511 EDDVRAQAARV-----RAAAGVVPVFIPEMGSRQHEALVLIYSAGAEMSMWLQNRPM 566
Qy 1177 QAGPESIGVPTAOEASVLRGTALGSVGSITGKIPSTRVPSDAITYRGSITGTAD 1236
Db 567 QA-PQGRNQCQRVYAPHGH--GSFTGSVTPPLPHIG--DAAAQOPWALPAPVAVAV 622
Qy 1237 VLYKGTITRIIGDSPSLDGRG--DLSLPKH--VYEGKGHVLSTYGMSVTQCS 1290
Db 623 AMSR-----RYDRQTKTFILQSLRRAYADMAVYGRADPAGCATVREAL-CA 668
```


QY 1075 FSYAPGHPRLPLGHDTPARVLPRLPTISNPPRLISSAKHPSVLERQIGATISQGMVOLT 1134
DB 2516 KRY-----RLIARFRHEGDEVDVEK-----DPYQIATTLVYDGLDE-----ISRETDLD-H 2561
QY 1135 VPYSEHAK-----APV-----GPTVMGLPLPMPDKLAPSGVQO 1169
DB 2562 APIADHANVYHHGDSRIEKNASPVATLEKKOKKVKSPKKEKVDVPAPPEK-----XE 2614
QY 1170 EQLSPRGAGP-PESLGVPTAOEASVLRGTALGVPG-----GSITKGIPTSRVPSDAIT 1224
DB 2615 IRLIARVTEPEAEIVDPVAVVETKEKKGDSKSLPTGFFSKTKTKTGYPETSE-A 2673
QY 1225 YRG-----SITGTPADVLVYKGTITIIIGEDSPSRIDR-----G 1258
DB 2674 YTGELDTDRJHDELGTSFEHEHNPAYSPKQIVTIAI-ETPAPEKKRYLLIARFRHEG 2731
QY 1259 REDSLPK-----GHVIEGKKGHVLSEYEGMSVTQCSK-----BDGRSSGPPHETAP 1307
DB 2732 DEDVVEKDDPYQIATTLVYDGLDEI-SRETDLDPRIADHANVYHHGDSRIEKN-ASP 2788
QY 1308 KRTYDMEGRV-----GRAISSASIEGLMGRAPPERHSPHH-----LKEQHIRGSI 1355
DB 2789 VATPEKKOKKVKSPKKEKVDVPAPPEKKEIRLIARVTEPEAEIVDPVAVVETKEKKGDS 2848
QY 1356 TQGR-----RSYVEAQEDY-----LREAKULKREGR-----PPPPSRD 1392
DB 2849 KSKLTPEGPFSTKYTTKTTGTGTPETSEAYTGELDTDRJHDELGTSFEHEHNPAYSPKQI 2908
QY 1393 LTEAVKTOALGPLK-----LKPANEG-----LVATVKEAGRSIHETPRE-ELRH 1435
DB 2909 VIETIETAPPEKKRYLLIARFRHEGDEVDVEKDDPYQIATTLVYDGLDEI-SRETDL 2967
QY 1436 TP-----ELPAPRLKEGSTTQGTPLKYDTGASTGSKKDVNSL 1476
DB 2968 TPRIADHANVYHHGDSRIEKNASPVATLEKKOKKVKSPKKEKVDVPAPPEKKEIRLI 3024
QY 1477 IGSPPRTPPRVAPRLVADARALERAQYEEESLKSPTASSGSGSANGAPVIVELKCP 1536
DB 3025 -ARVTEPEAEIVDPV-----VKETKEKKGDSKSK-----LPTGFF 3061
QY 1537 ROSPLVYEDHGAPFAGHLPKSGPVTMRRETPRLQEGSLSSSKASODKRLTSPR-REIAK 1594
DB 3062 SKTTTITTTGYP-----ETSEAYTGELDTDRJHDELGTSFEHEHNPAY 3106
QY 1595 SPHSTVPEHHPHPISPYEHNLKGVG-----GVDLY----- 1624
DB 3107 SPKQIVTIAIETPAPEKKRYLLIARFRHEGDEVDVEKDDPYQIATTLVYDGLDEISRET 3166
QY 1625 -RSHIPLAFDPTISIRGIRPLDAAAY-----LBRHLAPNPTYHLYPPIYLRGYPTA 1677
DB 3167 DLDHPIA-----DHANVYHHGDSRIEKNASP-----VA 3196
QY 1678 ALENQTIINDYITISQMHNTATMAQRAADMLRGISPRESSIALNVAAGPRG-IIDLSQ 1736
DB 3197 TLEKKOKKVK-----SPKKEKVDVPAPPEKKEIRLIARVTEPEAEIVDPV 3243
QY 1737 VPH-----LPVL-----VPPETGPATMDRLAYLPTAPQF-----SSRH 1772
DB 3244 VKETKEKKGDSKSLPTGFFSKTKTKTGYPETSEAYTGELDTDRJHDELGTSFEH 3303
QY 1773 SSSPLSPGGPTHLTPRTTSSSR-----EERDR-----R 1802
DB 3304 GEHPFEASSTISITLPAFDPSEHEKVIKKPSRFHMFSRMRHEGDEVDVEKPGYILISSELY 3363
QY 1803 ERDRREREKSLITSTTVEHAPIRWPGTEOGSSGSSGSGGSGSSSR----- 1850
DB 3364 EGPYNNMPPIGELTLPILHDSQVYH-----NGSGKLEKSSSEKRIKTPSPRVRPS 3417
QY 1851 -----PASHSH-----AHOSHSPISPTQ-----DALQOR-----PSVLANTGM 1883
DB 3418 EIKLVARIQPLEHVEDSEISEKLSPVKORSRAPSPFMSFMRKORGYPEISPLYGNDLV 3477

QY 1884 KGIITAVESKPTVLRASTSTSPVRAPATPPATPC----- 1920
DB 3478 TGRAAEVE-NTPIEVRASVPEHNPQOVSAQIPRESSQPRNFRLLIARFRHEGDEDEIEKPD 3536
QY 1921 -----LGGTLDGVYPTLMEPVLLPKAEARVARPERPRAD-----TGAFLA-----K 1962
DB 3537 AYKFVTEVEGRLDIRPV-----VELAQDPSIDHGOVYHHGDSMAEKNIK 3583
QY 1963 PRASGLEPASPSPKSGSEPRPLVPVPGHATIACTPAKNLAPHNASPP-----PAP----- 2015
DB 3584 PP-----KPVNTKKGLSPEDISEVHEIRLLTVASSSEEPEDVDVPAESTVPKPPRL 3638
QY 2016 -----ASADPPEKTKOSKP-----FSIQELERLSLGY-HGS-SYSP----- 2050
DB 3639 RFLSIGKKSPTSPSEPEKPGQYDGPVDLTSRDELEVMYLLGVSVPYPPKTPVKKV 3698
QY 2051 -----EGVEPV-----SPVSPSLT-----HDKG 2069
DB 3699 DKADEERKTYIVIRFRHEGDEPDADKPEAYGFSDDVYTGSLNEISRSSELEHAAIHDS 3758
QY 2070 LPKH-----LELDKSHLEBELPKO----- 2090
DB 3759 QVYHDGLSYKIDNKOKSKDRTELEPKTSRLVARVLPQOASDSSGSPKDDKGFRR 3818
QY 2091 -----PGPVKLGGEAHLPHL-----RPLPSPSPSSPLQIAPGVK 2127
DB 3819 FFSKASPKSGSPVPSPEEPIDVTGRVSELHENVPLQSRFPYSPKKS-----QEKPSK 3875
QY 2128 GHQVWTLAQH-ISEVITDDYTRHHPOQLSAPPLAPLYSPGAS-----CPVLDLR 2177
DB 3876 SSERTEBPVKYHLIARI-----RHEGDEETVENP-DTYGASTSYDGPLEETSKSVLVE 3928
QY 2178 RPPSDLYLPDPDHGAPARGSHSGGKSPSPNKTSVLGGEDGLEPSPPE-GMTPEG 2235
DB 3929 ETPISVSNVYVHSEISIRVEKORKAKISAPPEK-----DDQENVPSPSRPSFLRG 3981
QY 2236 HSRSAVYPLVYRDEQETSPSRMGSKSPGNTQPPAFSFKLTES-----NSAMVSKKOBIN 2291
DB 3982 OKRVLYSTPYISG-----PLETDRDLDENILPLSIPYHSSSQHAKRYLI 4029
QY 2292 KKLTHNNEPEYNIISQGTIEFMMPAITGTGLM-TYRSGAYOE-----HASTN 2339
DB 4030 ARIR-HAGDEDEVAAHKLHPVYGFPAQVYSGILEKTALCKEVDVPIGRFARVYHSGSS 4088
QY 2340 MGLEAIRKALMGKYDQWEESEPPLSANAFNLNASLSLPAIMPITADGRSD----- 2391
DB 4089 GGLVNNLKNNSPASKMDPEPKRRKQDQDKDKIKLARVLPKKTPEKSDREKATKE 4148
QY 2392 -----HTLTPGGGGAQVSGRPSRSRAKSPAPGLASGDRPPSVSVHSEGC 2439
DB 4149 EPLREAEIVRSTQSEAGQNIISTVGIYPSLTYDVP-----RPARIEGITYGDC 4197

RESULT 93
T20774
hypothetical protein ZK270.2d - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C/Accession: T20774; T27818
R/Kershaw, J.
submitted to the EMBL Data Library, November 1996
A/Reference number: Z19322
A/Accession: T20774
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-4667 <M1>
A/Cross-references: EMBL:Z81499; PIDN: CAB54224.1; GSPDB: GNO0019; CESP: ZK270.2d
R/Kershaw, J.
submitted to the EMBL Data Library, November 1996
A/Reference number: Z20423
A/Accession: T27818
A/Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-1603 <PAN>
A:Cross-references: EMBL:M92642; NID:g180757; PDB:AAA58427.1; PID:g180758
A:Experimental source: skin fibroblasts
R:Yamauchi, N.; Kimura, S.; McBride, O.W.; Horii, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J. Biochem. 112, 856-863, 1992
A:Title: Molecular cloning and partial characterization of a novel collagen chain, alpha A:Reference number: PQ0612; MUID:93203161; PMID:1284248
A:Accession: PQ0612
A:Molecule type: mRNA
A:Residues: 'GGR', 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1164, 'P', 1166-1603 <YAM>
A:Cross-references: GB:S57132; NID:9298641; PID:AA25797.1; PID:9298642
A:Experimental source: placenta
R:Kimura, S.
submitted to the EMBL Data Library, April 1989
A:Description: Partial nucleotide and amino acid sequence of a collagen-like protein from A:Reference number: S08012
A:Accession: S08012
A:Molecule type: mRNA
A:Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM', <KIM>
A:Cross-references: EMBL:X14963; NID:g29884; PID:CAA33085.1; PID:g930048
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C and subsequently O-glycosylated).
C:Genetics:
A:Gene: GDB:COL16A1
A:Cross-references: GDB:134045; OMIM:120326
A:Map position: 1p34-1p34
C:Complex: type XIV collagen may be a homotrimer, or a heterotrimer of two alpha 1(XVI) C:Function:
A:Description: structural component of extracellular fibrous polymer as a minor form product A:Note: may play a role in forming elastic connections at fibril surfaces
C:Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxylysine
F:1-31/Domain: signal sequence #status predicted <SIG>
F:122-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAT>
F:12-333/Domain: amino-terminal nonhelical #status predicted <NC11>
F:334-1577/Region: interrupted helical
F:334-360/Domain: collagenous COL0 #status predicted <COL0>
F:337-505/Domain: collagenous COL0 #status predicted <COL0>
F:521-554/Domain: collagenous COL8 #status predicted <COL8>
F:533-541/Region: cell attachment (R-G-D) motif
F:572-630/Domain: collagenous COL7 #status predicted <COL7>
F:652-722/Domain: collagenous COL6 #status predicted <COL6>
F:738-875/Domain: collagenous COL5 #status predicted <COL5>
F:887-938/Domain: collagenous COL4 #status predicted <COL4>
F:973-987/Domain: collagenous COL3 #status predicted <COL3>
F:1005-1007/Region: cell attachment (R-G-D) motif
F:1011-1432/Domain: collagenous COL2 #status predicted <COL2>
F:1266-1228/Region: cell attachment (R-G-D) motif
F:1472-1577/Domain: collagenous COL1 #status predicted <COL1>
F:1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NC01>
F:47,327/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 2.4%; Score 323.5; DB 2; Length 1603;
Best Local Similarity 21.4%; Pred. No. 0.00311;
Matches 376; Conservative 143; Mismatches 632; Indels 609; Gaps 99;

QY 751 IPSHTEAKNTGQNGKRP-----ATLGAD-----GPPPG----- 781
DB 179 VASVHVDSSASSQPLGRRPRPVGHVFLGDAEQKVSFDLQOVHIYCDPELVLEEG 238
QY 782 -----PPPPRRTRAP-----IEPTPASEAT-----GAPPPAPSPSPAPPVV 822
DB 239 CCEILPACCPPTSCARDTQSNELIINPQSEGVYTRCFLEPQMSYVDAQLTGIS 298
QY 823 PKEEK-----EEETAA--APVVEEGEEKRPAAEIADVGTGAEEVSECTEAEAGPAK 876
DB 299 OKAERGAKVHOBTADECPRCVHGARDSNVTLAPSGPGKGERGLPPRPGSGKEKG-AR 357
QY 877 GKDA-----EAEATAEAGLAEKEKSGRATTAKSGAPQDSSTCAADVDAAEG 931
DB 358 GNDGVRISPDAPLQCAEAP-KGEKESGA-----LGSGLPGSTG-----EKQOK 401
QY 932 GDKKRLSPRSLLTPTDPPANASPPK-----LDLKKQKORAAAIPIPIQVTKV 981

DB 402 GEK-----GDGGIKGVKPKGRDAPGEICVLIPKQKQDGPVGP----- 441
QY 982 HEPRBEDAATPKRAPRAPPPQNIQPSDAPQOGSSPRGSRBPAPADAEAAAEAKOK 1041
DB 442 -----GGLAGEPAPPLGPG-----PGLGLGTGG-DPGG-----DPGPGDGGSGIPKE 487
QY 1042 LPGDPCCWTSGLPAPVPPR--EVIRASPAAPPSAFSAPGHPPLPLGLHADRPAVLPRPP 1100
DB 488 GPGKRP-----GKPVKKEKGDPCVCFPLPBGFQNPGLPKPKPKG-----ERDDP 535
QY 1101 TISNPPPLISSAK-----HPSVLEHQIGALISQMSVQLHPVSEHAKAPVGVTLGLP-LP 1155
DB 536 VRARGDPQIGIKKEKGPCL-----SCSSVGAQ-HLVSTGASGVGPPGFLPLP 588
QY 1156 MDPKPLAPFSGVKQEQOLSPPGQAPPPESLQPTAQEASVLRGTLGSPGGSITRKIPST 1215
DB 589 -----GRAGVPGLKKEK-GNFEAGAPASPPGPPVGPAGIKG-----AKGEKCE 632
QY 1216 RVPDSAITVYGSITTHGPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEGKKG 1275
DB 633 PCPALSN-----LQDGDV-KVVALPGBSG-EKGEPPG--PPGFGL-PGKQG 672
QY 1276 HVLSEEGMSVYQCSKEDGRSSGPPHETAAPKTYDMEGRVGRAISSASIEGLMGRAI 1335
DB 673 K--AGEKGL-----KGQKGDAGNPGDPGTPGT-----GRPLGS-GEPPVQGPAG-- 714
QY 1336 PPEHSHPHHLKEQHINIGSTIQ-----GIRSYEAGEDYIARRAK-----LAKREGTP- 1384
DB 715 -PKGEKGDCTACPSLQGTIVDMAGRPQOPKQGEGRGKQKQPLPVGQGP 773
QY 1385 -----PPPPSRDLTEAYTKQALGPKLKAHGLVATVEAGRSIHIEPRELHPTPE 1438
DB 774 LKGVQGEPPGPKGVQ-----GP-QGEGAGGL-----PGLQGLP-----GPR 810
QY 1439 LPLAPRLKEGSIQTGTP-LKTYD-----GASTTGSKMDVRSILGSPGTPPPVHPLD 1491
DB 811 GP--PGPTGKRG-AQSGPVGVGATGVPGRASVSGPPGRD-----GQQGT----- 854
QY 1492 VMADARALERACYESLKSRRGTAASSSGGSIARCAPYIVPELGRKPSPLTYEDHGAPFA 1551
DB 855 -----GKRGTPGEGKPRG--EKGEPP--GEGSCPSQDGLIFS--GMFGA 891
QY 1552 GHL-----PRGSPVWMEP-----TPRLQESLSKSKASQ--DRKLTSTPREIAKS 1595
DB 892 PGLMWGSSWQGPQGPQGPICPPPPPPVPLGQVPGVGNGLPGQGLTALGLSLPIE----- 947
QY 1596 PHSVTPHHHPHPISEYHLIRGVSG--VDLYRSHIPLAFDPTSIPRGIPIDAAAAYULPR 1653
DB 948 -----QHLLKSIQGDVCQGGRAH----- 965
QY 1654 HLAHPPTYHLYPYLI-----RGYPDTALENRQTIINDYITTSQGMHNTATMAQR 1706
DB 966 -----PGYLVEKGEKGDQGIPIVGLDN-----CAQCFSLER-----PR 1000
QY 1707 ADMLRGLSPRESSLALNVAAGPRGILDSQVPHLPVLV-----PP-----TPGPATAM 1755
DB 1001 AEAARG-----DWSBDDPCGVSGPLGPPRGLPGQRGEGEPPPMKSGPPPPG-- 1047
QY 1756 DRLAYLTPAPQPFSSRHSSSPLSPG--GPTHLTYRPTTSSSERDRDRDRDREREKS 1813
DB 1048 -----PIGPFGPGA-VGSPGLPGLQGBRGLTGLTGDKG----- 1080
QY 1814 ILTSTTVEHAPYMRPTEGSSGSG-----SGGGGSSSRPASHAHQHSR 1862
DB 1081 -----EPGPQGPQGPATGPPGLGKIGERYGYSAGEKEPGRPGSEGLPGPPGP 1132
QY 1863 ISPTQ-----DALQORPSVLHNTGMGIITAVPSPKTYLSTSTSPVR-- 1908
DB 1133 AGPRGERPQGNSEKGDQGFQGPQGTGPTGSGPFGEKGVSPPPPPQAKSGSEGITRP 1192
QY 1909 -----PAATPPATHCPPLGGTLDGVPYPTIMEPVLL-----PKAPRYAARPPRAPDTGHA 1958
DB 1193 SGLPGSPQPPPPPPQIGPAG--LDGLDQKQKQPLGRBGPAGPPGLMGPPGFGFKTGHP 1250

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:39-98/Domain: von Willebrand factor type C repeat homology <WMC>
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 2.4%; Score 323.5; DB 2; Length 1497;

Best Local Similarity 20.5%; Pred. No. 0.0029;

Matches 379; Conservative 138; Mismatches 646; Indels 683; Gaps 95;

```
QY 716 ENVEAEALAHAGNEVPRGECGPRATVNNSSDT-----ESIPS 753
DB 71 DKIECPVLCANPITTPPEGCCPVCPQTGGDTSPGKRGKQKGERGLVAVVTGIRGRG 130
QY 754 PTEAAKD--TGONGPK---PPATIGADGP-----PGPPPTPTSTR-- 791
DB 131 PAGPGSGQPRDRGRKRGPRGRGPGIDGEPGMPQPGAPRPGHSHRPGDMSRPS 190
QY 792 -----APIETPASEATGAPPPRPAAPSAPPPVPE 825
DB 191 AQMAGLDEKSGIGSVGLMPGSGVPGVPRGPGVGLQGQCGAGPAGPGEPRGPGPGIG 250
QY 826 EKEEETAAAPVVEGEQKPPAAEELAVDTGAEE-----VKSECTEAAE 872
DB 251 SSGPGPPGPKPEBDEG---PGRNGTGVGSPGSGAGAPGAPGLPGLKGRHKGLE 306
QY 873 GPAKGDAAEAATAGALKAKEKGGSGRATTAKSSGAPQDSDSATCSADEVDEAEG 932
DB 307 GF-KG-----ETGARGA-----KEAGPTGMDA----- 329
QY 933 DKNRLSPRSLTPTGDP--RANSPQKPLDKOLKORAAIPIQVTKVEPPREDAA 990
DB 330 -----MGF---LGRGMPGERGRLLGP-----GAPQKRGA 356
QY 991 PTKRAPRAPPPQNLQPESDAQQRG-----SSPRGKSRSPAPRADKEAFPA 1037
DB 357 HCMPKRPGWGLGIPGSSGFPGNPMKGERGPHGARGEQGGQGETGPG---PA 411
QY 1038 EAQKLPGDPRCWTSGLPFVPVPREVIKASPHAPDPSAFYAPPGHPLPLGLDHTARVLP 1097
DB 412 GSQGLPG--AVGTG---TPGRKATGSAAGTSGRPGLA--GPPGSGPFG--- 454
QY 1098 RPPITSNPPLISSAKHPSV--LEROIQAISQ---GMSVOLHVPYSEHAKAPVG--PVTM 1150
DB 455 -----STGPGIRGQSGDPPGPRKGBAEPKGEPRPHIGQIPGPEGRKGRBPGDPTV 510
QY 1151 GLPLRDPKCLAPFSVCQKQEOLSPRQAPRPESLGVPTQAQASVLKGLTGVSPGSLTK 1210
DB 511 GPPGPMG-----ERGAP--GNRGPSPSDGLPGRKAQGERGVPSSGPKG--Q 555
QY 1211 GIPSTRVPDSAITYKGSITHTGPADVLYKGTITRI--IGEDS---PSRLD--RGRDEL 1263
DB 556 GPPGR--PGEPLPGAKRLT--GNPGVQGRBKLGPLGARGEDGRPPPSIGIRQPGSM 612
QY 1264 ----PKGHVYEGKKGHVLSYEGGMSVTQCSKEDGRSSGPPHETAPRTTYDMMEGRVG 1319
DB 613 GVPGRKSSGDLGKPG---EAG-----NAGVPGQRGAPGK---DGEVG 649
QY 1320 RA--ISSASIEGLMERAIPEPHSHHLEKHHRISITQGIPTASVYEAQEDYLREACT 1377
DB 650 PGGVPGPRGLAEBRGAAPP---GPTGF-----QGLP----- 678
QY 1378 LKREGTPPPPSR-----DLTEAYKTQALGPL-----KLKPAHEGLVATVKEAGRSIHE 1427
DB 679 -----GPPRPGEGGKADQGVPRGAVGLPGRGNGENGERG-----EPG--ITG 724
QY 1428 IPREELR--HTPELPLADRPLEKESI-----TQGTPLKTDYTGASTGSKKHIVSLI 1477
DB 725 LPEGKEMAGGHGDDGK--NPGRGTGITGTPRGGLQGMF--GBRGIAGTPGRKGD--RGSI 780
QY 1478 GSPGRFPFVPHPLDVAADARALERACVYESLSKSRPTASSGGSITARKGAPVLPBELGK- 1536
DB 781 GERK--AGTAGNNGARQLPPLPAPPPAGLGLAP-----GEBG 817
QY 1537 -----RQSLTYEDHG---APFGHLPRGS---PVTMRBETPPLQSGSLSSSKA 1579
```

```
DB 818 PRGLVPGPSRQNPQSGRNGNPTGAVFAVG--PQSSDQPGVKSGPGEPRGQGDAGS--- 872
QY 1580 SODRKLTPRELIASPHSYTEPHHPHPISPENHLLRVSQVDLYRSHIPLAFD--PTSLP 1638
DB 873 -----PGPQGLASP---GPHGPHAV--PGLKSGRGTGQF-----PGATYPPGSAK 913
QY 1639 RGIPLDAAAYVLPRLHLPNPTYPPLVP---PYLIRGVPDTALLENROTIIINDYITSQQ 1694
DB 914 RVGPPRPG---APGAPRAGEPQKEPRPLRQDPG----- 947
QY 1695 MHNTATAMAQRADMLRGLSPRESSLALNYAAGPGIIDLQVPHLPVLVPTPGTPATA 1754
DB 948 -HGVGD-----RG-----PAGP-----PGSPGDK 966
QY 1755 MDRLAVLPAP---QPFSSRHSSPPLSPGQTHLTKPTTSSSRERDRDRERDRNR 1810
DB 967 GD-----PGEEDQP-----GPDGPPGAGTGGRGIVGMDQR----- 999
QY 1811 EKSLTSTTVVHAHPIWPRGTEQSSGSSGSGGG--GSSSRPASHAHQHSPTSPRTQ 1868
DB 1000 -----GYTMGRLP--GRAGTGKVGTPGATGDKPRPGVPGSGNGPVGEPGPPGAPN 1052
QY 1869 DALQQRPSVLTNTGKGIITAVEPSKPTVLRSTSSVPRPATPPATHCPLGGLDGV 1928
DB 1053 DGTPRDGAVERGDRG---DPC--PAGL-----PGSQGAPGPGVGAAPDAG 1096
QY 1929 Y---PTLMEPVLLPEEAR--VAREPRADTGHAFIAKPPARSLERASSSKSEPRP 1983
DB 1097 ORGEPGSRPGVPVPPRAGKRLPGQGRGDKG-----DNDRDRDRQ-- 1139
QY 1984 LVPVPSGHATTARTPAKULAPRHASDPDPAP--PASASDPREKTQSPFSIQLELRS 2040
DB 1140 -----KCH-----RGTGLQGLPGRPPGBOGSAGIIPPFRPGRP----- 1176
QY 2041 LGTHSSYSPEGEVBPVSPPSLTHDKGLPRHLELDKSHLEGELRPKQRPVYLGEA 2100
DB 1177 -GPVSSSGKEGNRGLPGLIPGPGVAVSG-----EAGREGP--PGEPPRPPGPP 1224
QY 2101 AMLP-----HL-----RPLP---ESQSSPPLQTPAPGVKQGRVYTLAQHISEVT 2144
DB 1225 GHILTAALGDIMGHYDENMPPDLPEFTEDQAAADDNTKDPGI--HYTLKSLSSQLETKMS 1282
QY 2145 ODYTRHHPQOLSAPLPAPLYSPGASCPLDLRRPPSDLYLPPPHGAPARASPHSEGG 2204
DB 1283 PDGSKKHPR-----TCDDLYKCHT-----KQSGE 1308
QY 2205 RSPERNKTSVLGGEGDIE-----IVSPEP---GMTEPGHSRAVPLY 2246
DB 1309 YWIDPNQ-----GSADDAIKVYCNMETGETCISANPASVPRKTWMAKSPDNKPVWYGLDM 1364
QY 2247 RQGEQTEPSRMKSKSPGNTSOPPAFFSKLTENS-----AMVSKKQELINKLKN 2295
DB 1365 NRGQF--TYGDYQSPNTAITQMTFFRLSKASQNLTYICNTVGYDDQAKMLKAVV 1422
QY 2296 THNRNEPEYNIQPGTEIFNMPAITGTGLMTRSOAVOEHASTNMG 2341
DB 1423 LKGSNDLE-----IKGBGNIRRYTVLDQTCRKRG 1453
```

RESULT 91
S23810
collagen alpha 1 (XVI) chain precursor - human
M:Alternate names: procollagen alpha 1(XVI) chain
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence revision 28-Oct-1994 #text_change 15-Sep-2003
C:Accession: S23810; P00612; S08012
R:Pan. T.C.; Zhan. R.Z.; Mattei. M.G.; Timpl. R.; Chu. M.L.
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992
A:Title: Cloning and chromosomal location of human alpha1 (XVI) collagen.
A:Reference number: S23810; MUID:9235339; PMID:1631157
A:Accession: S23810
A:Molecule type: mRNA

QY 1562 MREPTPLRQESLSSSKASODRKLSTREIAKSPHSTVPEHHHPHPIPYEHLRGVGV 1621
D 1335 -----RTR-----TTSKPSQSLKIAEKGDALSLNLT-----EQHQPRLDAQELLEKEISEQ 1361
QY 1622 DLYRSHIPLAFDPTSPRGIPRLDAAAYLLPRHLPANPTYPHLYPRLIRGYPDTALEN 1681
D 1382 D-----GSDQSF-RTPRMD-----LKPQODEBEL-----ESLGTSSANQD 1416
QY 1682 RQTIIND-----YITSQMHNTATITMQRADMLKGLSP-----RESSLALNTAAG 1727
D 1417 CPQGVKPRMOMOVOLRRQNHSLIEDLRRLPQLQAEHLBNQDVERDRIRIGIATLLSAS 1476
QY 1728 PRGIIDSQVPLVPLVPTPTGTATAMDRLAYLPTAPQPSRHSPLSP----- 1779
D 1477 HQDCQTSSEEPQSVQVPESP-----AATER-----OEPLSPQSDGPVGPRAIKKGL 1525
QY 1780 ---GGPHTLTKPTTSSSERERDR--DREDRDREREKSLITSTTV-EHAPIWRPTEQ 1833
D 1526 SRQGRREGQKPLRGFRGADPOVDNVGSHVEXKSKILDQERLDQHHGIMKPSBEQ 1585
QY 1834 SSSGS-----SSSGGGG-----SSRA--SHSHAHQHSPISTPTQAL 1871
D 1586 PSSASDLPCSSDPSLSVNDGRSIRDSGPLKPTSGRSPFRPYRRPOERSQSHMAQAC 1645
QY 1872 QORPSVLHNTGKGIITAVEPSKPTVLKSTSTSPVRPAATPPATHCPLGGLDGVYPT 1931
D 1646 RRCSEADHHPDSDSLIDM-----IDTSNRSRPMPP----- 1677
QY 1932 LMEPVLPKEAPRVARPERPRADTGAFLAKP-PARSGLEPASPSKSEP 1961
D 1678 -----LPEEKPMALQAHQKLAAYHAFRCPLPTAISRLP--RRQNTPEP 1720

RESULT 88

T48818
giucen1,4-alpha-glucosidase related protein [imported] - Neurospora crassa
N:Alternate names: protein 68B2.20
C:Species: Neurospora crassa
C:Dates: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T48818
R:Schulte, U.; Altm, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224541
A:Accession: T48818
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2022 <SCH>
A:Cross-references: EMBL:ALJ53821; GSPDB:GN00112; NCSP:68B2.20
A:Experimental source: cosmid contig 68B2; strain 74
C:Genetics:
A:Gene: NCSP:68B2.20
A:Map position: 2
A:introns: 1192/3

Query Match 2.5%; Score 324.5; DB 2; Length 2022;

Best Local Similarity 18.7%; Pred. No. 0.00037;

Matches 430; Conservative 274; Mismatches 892; Indels 701; Gaps 102;

QY 533 EEKPEVENDKEDLLKTKTDITSGEDNDEKEAV--ASKGRKTAQSGRRKGRITSMANE 589
D 62 KKKPELKKPEBLT---GDFLLGRPSQORVAQASAKRKT-----MSMD 104
QY 550 ANSEALITPOQSALASHELNSSRWTEBEMETAKGLLEHGRMSATAAM----- 640
D 105 AONVREELRQERAAVAVKLQOPG-----GVNDRYKAWOKASQAQVAKGLPT 152
QY 641 -VGSKTIVQCKNFYNYKKRQNLDEILOHKLME--KERNARKKKKAPAAASEEAAPP 698
D 153 FVAEABARBEPTFAVNLG---DDVDEDRMRITMRQKRPQKKPKT----- 196
QY 699 VVEDEEMASGVSGNEEWEAEALHASGNEVPVPGECSPATVNNSSDTESIP----- 752
D 197 -----ETVBPKGTSGSDSTAKEKNKVLGRMADDRP-----DVLIEKARPRPLPKKRIIS 246

QY 753 SPHTAAKDGONGPKPPATLGADGPPPPPTPE---RRTSRAP-IEPTPASEATGATP 808
D 247 DDHMMKQOQKGLAKSPV-----PEAGQPIPDFLKRTAQNSVQKKIRDAQORELP 300
QY 809 PPAPSPSAPPVVPKEKEKETAAAPVEE---GEOKPRAAEIADVGAKEEVPK 863
D 301 PPPPVARPTVTKYRIAKGETVTEVEDEDDHNRGALSIPDVRPRKRSSSGSSRSS 360
QY 864 SECTEEA-----EEG-----PAKQDAEAAEATBEGALKYKEKEGG 899
D 361 NPRTQCCPNDGIRVRSLOKKSNDGIRISPVPARSLPDGQIKVPGPVASDSSRS 420
QY 900 SGRATTKSSG-APQDSSATCSADYDAEAGDKNRILSPR-SLTPFGDRAVAP 957
D 421 ISTVPSSSSGRTSDSGSKT-----PPGASPPRRASTP--PRRSTP 464
QY 958 QKPLDLKQKOR---AAATPIQVTKNHEPRDAAPTKEAPP-----APPPONLQ 1007
D 465 LRKASTKPRARSDHSAASDDVIEV--IYEESEVSXKSPSPPKRLRPPPKRLP 522
QY 1008 ESDAPQOPGSSPRGKSSPAPPADKEAPAAEQKLPQDPCTWISGLPPPV-PEEVIKAS 1066
D 523 RRRG-SRQGARKEKRSRSPPTT---AQTEITTDPR--RGADKPMPTPRNNGSS 574
QY 1067 PHADPSAFSVAAPPGHPLPLGLHDTARVLPRLPTISNPPLISSAKHPSVLERQIGALS 1126
D 575 GEDSD-----RRPTAAGIDL-----AEIP 595
QY 1127 QGMSV-QLHVP-----YSEHAKAPVGPVTGLPLPM--DPKCLAP-----FGVKQEQ 1171
D 596 FGSAFSELEPLRGLHGTQRTAARQAKP---KKPQONESLKVLPVNLKKVMTGA-MEK 650
QY 1172 LSPRGQAGPERSLGVPTAQAESVLRGTALSGVPGSGITKGIPTSRVPSDAILTRGSLTH 1231
D 651 MQEAAEPPRPPTQKAPRIEISWLNNTVDPVEEG--MPLNPVVAEPLRVS----- 700
QY 1232 GTPADVLYKGTITLIGEDSPSLDRGREDLPRGHVYIEBKKGHVLSEGGMSVQCSK 1291
D 701 -TPE--RNSKEKLVDRDLPAHREKAPERN-----SK 728
QY 1292 EDGSSSGPPHETAPKRTYDMEGRVGARISASIEGLMGRAIPREHSPHNLKEQHNI 1351
D 729 EKLVDITVSTHKATPE-----ISKXASAEIAPRKERPQPVTHIELRKST 775
QY 1352 RGSITQGIPRSVEAQBEDYLRE-----AKLLKEGTTPPP-----PSSRD 1392
D 776 RSDSEVTPK-----RRKSPTTTPPSAAGIKRSTPSTRSPTLPSSSGSVKQPRD 825
QY 1393 -LTEAYKQNALGPLKMP-AHEGLVATYKAGRSIHEIPRELIHTPELPLAPPLKES 1450
D 826 VLKEAFKQES-SANKIAPMVTPSCETDVESEPETEHD--ESRRSPQOQSPDSYKRS 881
QY 1451 ITQGTPLKYDTGASTGSKKHVRSLSGSPKRTPPPVPLDVMDARALERACYEESLKS 1510
D 882 ASPRPSRADSYSSSEPTR-----GFSRRRRPTSDLDLSTLISEDREKQEPYXK 933
QY 1511 RPTGASSGSGIANGAPVIVELGKPROSLITYEDHGAPFGHLPKSGPVTMREPTP-- 1568
D 934 KD---TDSVSTVSGSTVQAPAEBAFGQPTPLTREASQOQSQTQISRSKSGSLKRLITTHS 990
QY 1569 -----LOEGSSSKASQDRKLT--STRPEIKSPHSTV-----DEHHPHPISP 1610
D 991 DLVSVLSLPDDGQLVPPERSKSTASRSLHKKPSKANDSRVNDLLEFADENHFYH--RE 1048
QY 1611 YEHLRGVGV-----DLYRSHIPLAFDPTSPRGIPRLDAAAYV--LPRH-- 1654
D 1049 LKTIYDGVVPLVLFNEFVHGDVVDADARKTDSMA--KAVVNVGVALLEKLMYHKAPLHDI 1106
QY 1655 -----LAPNTYPLHLYPRLIRGYPRTALLENKQTIINDYITISQMHNTATAMORA 1707
D 1107 RRLLEMLEAVSPV-----NNYLDVWRLGQODILVNLAPSPGKIDENDS----- 1150

QY 1303 ETAAKRTTYDMMEGRVGRALSSASIEGLMGRALPPERHSPHILKEQHILRGSTIQGIPRS 1362
DB 1101 GTPGP-----AGLATTLP-----LTTRAAQGRAPALSSS 1133
QY 1363 YVEAOEDYLREAKLKEGTTPPPPPSRDLTEAYKTQALGPLKPAHEGLVATVKEAG 1422
DB 1134 WQPPAN--MNRPEPSCRDTFAPTHALSSPA---EADGSVAVPQAGVARIPIPR 1188
QY 1423 RSIHEIPREELRHTPELPLAPPLKSGSTQGTPLKYDTGASTTSSKKHVAISLIGFGR 1482
DB 1189 TSSHADPE-----ABPPWSGRLPAFGVLPAT-----EPRGTPGSPSG 1227
QY 1483 TEPVPHPLDVM-----ADRAL-----BRACYEESLKSPGTASSS---G 1519
DB 1228 TQEPGPGLEKPLRQPPPEKALDLEKPLRQPPPEKALDGLISOGSALATQOMLG 1287
QY 1520 GSIARGAPVIVPELGR--PROSP 1540
DB 1288 GQ--RG--VRVPLLSRLPYQDP 1306

RESULT 87
T26135
hypothetical protein Y6B3B.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 29-Oct-1999
C:Accession: T26135; T27328
R:DoBeon, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20158
A:Accession: T26135
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1829 <W12>
A:Cross-References: EMBL:Z82068; PIDN:CA04901.1; GSPDB:GN00019; CESP:Y6B3B.1
A:Experimental source: clone W04A4
R:White, S.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20345
A:Accession: T27328
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1829 <W12>
A:Cross-References: EMBL:AL032655; PIDN:CAA21727.1; GSPDB:GN00019; CESP:Y6B3B.1
A:Experimental source: clone Y6B3B
C:Genetics:
A:Gene: CESP:Y6B3B.1
A:Map position: 1
A:Introns: 20/2; 53/3; 82/3; 329/3; 332/3; 355/3; 396/3; 462/3; 1799/3

Query Match 2.5%; Score 325; DB 2; Length 1829;
Best Local Similarity 19.2%; Pred. No. 0.00032;
Matches 356; Conservative 221; Mismatches 616; Indels 659; Gaps 88;

QY 370 HEVSEIIDLSECEENLEKQMOLAVIPMLYDADOORIKFI-----MANGMLADPM 420
DB 289 HEIAGOLP-WARQEAISKLEKLG-----YDA--RIHEIAGNSAIDQASNDIQAYQA 339
QY 421 KYVKDQVNNMWSDEKETFRKFMQHPKPNFGLISFLERKTVACVLYYYLTKKNENYK 480
DB 340 KI-----QILE--EPVEDQTSKKVLLQDPYK-----DQKFGESVQOVVLEKIGESSEK 385
QY 481 SILVRSYRRGKSQQQQQQQQQQQQQQQPMF--SSQSEKDEKEKEKAEE--EKEKE 537
DB 386 -----DQKIEDOVLQENLELEFPGRPAVEEDPKQVVEGISKNRVEEDE 432
QY 538 VE--NDKEDL-----LKEKTDTSGEENDKEAASQGRKTAASQGRKRITRSM-- 586
DB 433 VELIHEDULTIATPLSHMDVTENIONLDANAAEPLEPFLAAVVEQIAKDKILEEYQI 492
QY 587 -----ANEANS-----EAATP--QQSABL-----ASMEINNESS 613
DB 493 QDDMHPQNDALSESDIMEAREAQTVQDFIEDPVTEVEVEETELPEDRKDPALHLENNH 552

QY 614 RTEEMETAKKGLLEHGRNMSATAPMVGSKTVGCKNFYFNKKQNLDELLOOHKLK- 672
DB 553 DPTADVEDPSVDVFD-----LVDS-----DALEKVELKD 584
QY 673 ---MEKERNARRKKKKAPAAASEAAPVVEDEMEASGVSGNEEVEAEALHASGN 729
DB 585 QKVSQETSASDQKK-----SEADKDLSGEGSEBDQAPELGKGDSEVESKDPVABAN 637
QY 730 EVPRGEGCGPATVNNNSDTESIPEPHTEAADTQONGKPPATIGA--DGPPG-----PP 783
DB 638 LIP-----SVPRSGDLQATDSKESHSDKES-----KDPQVLDANNDRPDQSESDP 684
QY 784 TP-----PRTS-----RAPLETPASEATGAPTPAPSPSAP 819
DB 685 SPLESKHQVILKATLPPREGSVESKNHOQAAGVHPHAGSESGKTEDPNVLESPPAPK 744
QY 820 -PVVPKEKEEET-----AAPV-----EGEERQKPPA----- 848
DB 745 NPVEHNLEAQNPDPPIADPAPPIRIKRYKVKVPAVGVYKTRTQKMKLEDRD 804
QY 849 ---BELAVDTGKAEPEPVKSECTEAEF-----GP-----AKGKAEMAEATTEGLK 892
DB 805 RTSMEEVVAKTEKKDQPPASDLSDSRELSVAAGPQAPKRRRRGRMEDEE-TARKRK 863
QY 893 AEKKGCGSGRATTAKSGAPQDSDSATCSADEVDEARGDKNRLSPRPILPTGDP 952
DB 864 GRKLE-----ASDPRDQSGAPGCAPEPVKKTWSRKRYVDSENV-----PQ 907
QY 953 ANASPKPLDLKOLKORAAAIPIQVTVKHEPPREDAAPTRAPAPPPQNLQPESDAP 1012
DB 908 MNTRRK---ISEREYPAELGVPARRKRKDPVEDAG-----AQ 944
QY 1013 QQPGSSPR-----GKRSRAPPADKAFPAEAKL-----PQDPCCWTS 1051
DB 945 KDQGSATWVCIKIRFDGIPKIRDPA-----KSGAEDDELNFTSSDVLQCGSGVCS 999
QY 1052 GLPFPVPP-----REVIKASPHAPDPSA-----FSVAPPGLPLGLHDTARPVLP 1097
DB 1000 ---PATRSTSEQDSTQSGEDSDTSKIPCFEVSMP-----ISEPIL- 1042
QY 1098 RPTTISNPPPLISSAKHS-----VLARQIGAI--SQGMSVOLHVPYSEHAKAPV-G 1146
DB 1043 -EALIKNDPILIFNAGSGSSASASSSEAGPRIGIYEDQFOUMIGNIPGDDSGSVQG 1101
QY 1147 PYTWGLPIPMOPKTLAPFSGVKGQEL-----SPRQAGPESLGVPTAQEASVLGTAGS 1202
DB 1102 SETNKPPEPKR-ELQDVAGTDQTRILLATSPDSEA---SANDHAKDPMT--ETTRK 1154
QY 1203 VPGSITKGIPTSTRVPSDAITVRSITHTGPADVLVYKTYTRILIGEDSPRLDRGRDS 1262
DB 1155 LEAEKPOQOFKVVYFKQKLE-----SEBDGNGKT-----DPNQDL 1193
QY 1263 LPKGHVIEGKKGHVLSYEGMSVTQSKEDGRSSSGPHETAPKRTTYDMMEGRVGRAL 1322
DB 1194 M-----LYKGSREHIDL-----ELIOEKASYQSENNM--EI 1225
QY 1323 SSASIEGLMGRALPPERHSPHILKEQHILRGSTIQGIPRSVVEAOEDYLREAKLKE 1381
DB 1226 AOPSVSSEMSDLP--GPHVLKEQRDTSSEF-----AAYLADLKKIYKN 1271
QY 1382 GTPPPPPSRDLTEAYKTQALGPLKPAHEGLVATVKEAGRSIHEIPREELRHTPELPL 1441
DB 1272 ISIAEQSGSDSTN-----PNSDKOC----- 1294
QY 1442 APRPLKESITQGTPLKYDTGASTTSSKKHVAISLIGSPGTPEPVHPLDVMAARALER 1501
DB 1295 -----RKKREL-----LPQLDRSQIA----- 1311
QY 1502 ACYESLSRPGRTASSSGSITARGAPVIVPELGRKROSPLYEDHGAFAAGLPRGSPVT 1561
DB 1312 ---REDLK-----PKISTPEES---ED-----PARSOVT 1334

QY 2029 KPFSTGELSLSTGSSSSYSPGVSPVSSPLTHDKGLPKHLEELDKSHLEGELRP 2088
DB 1029 -----GFTSTGAP---GPIGPTGARGST-----GPAGP 1054
QY 2089 KOP-GPVKAGE---AALPH--LRPLPESQSSSPLQTAGVGHQVVTLAQHISEV 2142
DB 1055 SGPSPAEBERGRTGAGKHGHPVSGGLGLOCTSGPMGE--PGANGEO----- 1100
QY 2143 ITQDYTRHHPOQLSAPLPAPLYSFPGASCPLVDLRRPPSPLYLPPDHGAPARSGPSSEG 2202
DB 1101 -GQOQSTR-----GLPGA-----RGSNGNDGPGSGPR 1124
QY 2203 GKRAGE-NKTSVVLGGSGDGIETPVSPPEGMTPEGHSGSAVYPLLYRDEQETPS-RMGSK 2260
DB 1125 GPDPGEGRGPRGSGSGSGPPGPPGPPGPGVQSS-YGVR-----PSFGSGK 1176
QY 2261 SPGNTSQPPAFSSKLTESNSAMVSKQEIINKKLTNRNREPEVNIISOPTEIFMMPA1T 2320
DB 1177 GQSSS---PYGAYRDDSKNDVAKIQDTL-----LGATIS 1208
QY 2321 GTGLMTYRSQAVQEHASTN 2339
DB 1209 ALGQOIELIKAPQCKAKTN 1227
RESULT 86
T09219
basal transcription factor SNAPc large chain SNAP190 - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09219
R:Wong, M.W.; Henry, R.W.; Ma, B.; Kobayashi, R.; Klages, N.; Matchias, P.; Strubin, M.;
Mol. Cell. Biol. 18, 368-377, 1998
A:Title: The large subunit of basal transcription factor SNAPc is a Myb domain protein b
A:Reference number: 216616; MUID:98078693; PMID:9418884
A:Accession: T09219
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residue: 1-1469 <WON>
A:Cross-references: EMBL:AF032387; NID:G2641556; PIDN:AAC02972.1; PID:G2641557
A:Experimental source: tissue type fetal cell teratocarcinoma
C:Genetics:
A:Gene: SNAP190
C:Function:
A:Description: transcription factor; required for transcription of snRNA genes
C:Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat homolo
C:Keywords: DNA binding; RNA biosynthesis; transcription factor
F:396-447/Domain: myb DNA-binding repeat homology <MYB>
Query Match 2.5%; Score 326.5; DB 2; Length 1469;
Best Local Similarity 20.1%; Pred. No. 0.00022;
Matches 331; Conservative 196; Mismatches 573; Indels 543; Gaps 76;
QY 84 LRPEHSYLPETGKSEMERIESKRPRLLEPLLRPS-PLLAT-----GQPGSEDLTRD 138
DB 21 LDPGSGSGHVEISSESL-----SDEADSLPSEDLPADPPISEERMGESANDDEDKRD 76
QY 139 RSLTGKLEPVPSPPHDPELELVPPRLSKEELIQNMDRVDRITMVEQOISLKKKQO 198
DB 77 KTL-----PDDPETCLQNMVYQVIOEKLA-----ANLLAQNRQOE 116
QY 199 QLEENAA--KPPPEKPVSPPIESKHSLSVOITIDENRKAEAARHLEIGAPQVEPL 256
DB 117 ELMRLASGKGTVDGKSLPSTYMGHMKPYFDK-----VTGVPAN-- 162
QY 257 YNQPSTDRQ-YHENIKINQAMRKLLIVYKRNHA--RKQKQKFCQRYDQ---LMEAL 309
DB 163 -----EDTEKKAQGIK--AFEEILVTKKMKMEKALKRSVSDRLQRLQKLLKLEVL 215
QY 310 EKKVERIENPPRRRAKESKREYKQPEIRKQRE--LQRMQSRVQSGSGLSMSNA 366
DB 216 HOKQKVSSELRQLEKQGR-AEKELIDINQLPBEALLGNRLSDHMEKISINIFES 274

QY 367 RSEHVEEIIDGSEQENLEKQMSQLAVIPMLYDADQQRKIFINMGMLMADPMKYKDR 426
DB 275 RSAELIKFMQW-SHHSINKQ----- 295
QY 427 QVNMMSQOEKETEREKFMQHPKNFGLIASLEKRTVAECV-----LYVLTKNENYKS 481
DB 296 -----EWREBERERQALIAAH-----GHLEMQKIAELGTSRSAFOCLOKQOQNHKA 343
QY 482 LVRSYTRRGKSGQQQQQQQQQQQQQQPPRBS-----QEKDE-----KEKEEA 529
DB 344 LKRKEW-----TEEDRMLTQLVQEMRVGSHLYRRIYIYWEKGRSMOLLYKWTSLDPLG 399
QY 530 EKEKEKEVEVNDKEDLLEKTDGTSGEQNDKEKVAASGKRTAASQGRKRIRTRSMANE 589
DB 400 KKGWAPR-----EDAKILQVAKYGEQDWKITEEYVGRSDACDRYLRLHPSL--- 451
QY 590 ANSEBAITPQOASAEIASMELNESSRWTEEMETAKGLLEHG-RNWSAIARVSGSKTVSQ 648
DB 452 -----KGRMMLKEBEQILIELIKYGVGHMAKIASLPHRSGSQ 490
QY 649 CKNFYFNKKQNDLELLOHKLMEKERNARKKKKAAPAAASEBAAPPVVEBEWEAS 708
DB 491 C-----LSKWKIMGKKQGLRRRRRRRARS-----VRWSTSS 524
QY 709 GVSGNBEEMVEAEALHASGNEVPRGSGPATVNSSDTSSIPSPHTEAKDTGQNGPK 768
DB 525 GSSS-----GSGSGSSSSSSSEDEP-----EQAQAGEG 555
QY 769 PPATLGADGPPGPP--TPPRRTSRAPLEPTPASEATGAPTPPAPSPAPPVVPKKE 826
DB 556 DRALLSPQVWVPMQDLMWPARGSTQPMR--GGAQAMIGGPAASLSPRGSS--ASQGS 611
QY 827 KEETAAAPVYEBEBOQKPPAEELAVDTGAEEPV--KSECTEAEERKPKGDAEAE 884
DB 612 KEASTTAAAPBEETSPOVPA-----RAHPVPVPSAASASADTRPA-GAEKQALE 661
QY 885 A-----TARGALKAEEKEGSGRATYAKXS--GAPDSSSATCSADEVDEA--- 929
DB 662 GGRLLIVPVETVLRLVRLANT-----AARCTQKQLOKRPPLPTSPGVSDDSTARKSHVQ 717
QY 930 -----EGGDK-----NRLSPRPSL-LTP-TGD--PRANASPOKPLD----- 963
DB 718 WLBRHATQSGRRMRHMLHRLLRLRLLATVPMWGDVVVCTQAS-QRAVVGQOAGDL 776
QY 964 -KOLKQRAAALP-----IQVTKHEPRBEDA--PTKAPRPPPPQULQPSD 1010
DB 777 REOLQOARLSTPVFTLFTQLFHIDTAGCLEVVERKALPRLPQAGARDPVHLLQASS 836
QY 1011 APQPGSPRGRKSRSPAPPAKFAAQAQKLPQDPCMTSGLPFP--VPREVTKASPH- 1068
DB 837 SAQS-----TGHLFPVNPQAEASKASAKH 861
QY 1069 -----APDSAFYAPPGHPLPGLHDTAPVLPRPPTISNPPLI----- 1109
DB 862 GSRRLASRVERTLPQASLSTGP-----RPKQTVSE--LLOEKELQ 903
QY 1110 -----SSAKHPSVYERQIGAISQMSVQIHPYSEHAKAPGPVTMGLPL--PMDPKLA 1162
DB 904 EBARAEATRGVVLVPSCL-LVSSSVILQPLPHPHRPHGAPAPVLANVPLSGFAPAAK 962
QY 1163 P-----FSGVKQEOISPRGQAP-----PSLIG-VTPAQOESVY-----R 1196
DB 963 PGTSGSQWQAGISADKSLSTM-QALPLAPVPSHAEAGIAPASQAPALGPGQISVCSBS 1021
QY 1197 GTALGVPGGSITKGIPST-----RVPSDSAITYRG-SITH-----GTPADVLYKGT 1242
DB 1022 GIGQSGAAPAARSKQGLPAPFPPLPGAPSPPLPVQPLSLTHIGBHVATSVPLPVTWLT 1081
QY 1243 ITRIIIGEDSPRLDRGREDSLPKGHVYIEGKKGHVLTSEGGMSVYQCKEKEDRSSSGPH 1302
DB 1082 AQGLLPVVPVAVV-----SLPR-----PA 1100

```

Db      953 WHISLVNVPFTLTGQKELLGRTGLVVEAFWIDIKSHCFYVYSTEVEAVATRTALHGVK 1012
      1475 SLISGPGRTFPVPLVDWADARALERACYEES-LKSRPGTASSGGSSIAAGAVIYVEL 1533
      1013 -----WPOSNPKFICADYABODELDYHNGLLVDRPSLTKVEEQIIPPLHPPPP 1063
      1534 GKPROSPLT-YEDGAPPAAGHLPGSGPYTWMEPT-----PRLQSGSSSSKASQDR 1583
      1064 VPPQGPRAEOROEBAVAEQWAEERERERERERERERERERERERERERERERERER 1122
      1584 KLTSTPRELAKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTSLPRGILP 1643
      1123 R-----RKERAKSEKSEKSEKAKOEPPAKLLD-----DLFRK-----TK 1158
      1644 DAAAAVYLRHLAPNTTYPHLYPPLIRGYPDTALENNQTIINDYITSQQMHNHTATAM 1703
      1159 AAPCIYMLP-----LTDSDIVQKEAEBAERAKERERKKEQEEQEEQEEQEEQEE 1207
      1704 AORADMLRGLSPRESSLALNVAAGPRGIDLSQVPHLPVLVPPPTGTATAMDRLAYLPT 1763
      1208 RENNRQLEERKREHS-----REDR-----REDR-----1228
      1764 APOFSSRHSSSPSPSPGFTLTKPTTSSSERERDR-ERDRERERK 1812
      1229 -----ERERERERDRDRDRERDR 1249

```

RESULT 85

```

S23809 collagen alpha 2(I) chain homolog - sea urchin (Strongylocentrotus purpuratus)
C/Species: Strongylocentrotus purpuratus (purple urchin)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
A/Accession: S23809
R/Expósito, J.Y.; d'Allesio, M.; Solursh, M.; Ramirez, F.
J. Biol. Chem. 267, 15559-15562, 1992
A/Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) col1
A/Reference number: S23809; MUID:92348411; PMID:1639795
A/Accession: S23809
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1414 <EXP>
A/Cross-references: EMBL:M2040; NID:g161435; PID:AAA30035.1; PID:g161436
C/Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C/Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:107-1414/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```

```

Query Match      2.5% Score 326.5; DB 1; Length 1414;
Best Local Similarity 20.2%; Pred. No. 0.00021;
Matches 356; Conservative 110; Mismatches 576; Indels 717; Gaps 85;

      729 NEVPRGCS-----GPAIVNNSDTEISIPSPHTEAKDTQNGPKPATGADGP----- 779
      38 SEGRGRGKCGKGEFGDADINSANFPGLPGP-----VGPSPGSPGSPA--GNNGPDPNG 91
      780 -----PGPPTP-----RTSRAPIEPT-----PASEATGATP 808
      92 PRGNPGMDGLTGLRIGIPGPPEPKSGSLVASAQTSFNKPSLAGYQYQDAAGTGP 151
      809 --PPAPSPAP-----PVVPEKEKEBETAAPVEGESEQKPAEELAVDTGKAEPV 862
      152 RGPFGPESRGPQGLTGPSSGSGTGPSSNGSPGSGLPGRPS-----DQDDGTP- 203
      863 KSECTEAEAGPAKDAEAALATBAGALKAEKKGSGRATTAKSGAP--QDSDSAT 920
      204 -----GSQQRGPA-----GTPGSRGTGMGAPGMKGHQLPG 237
      921 CSADVDADAGDKNRLSP-----RPSILTPGP--RANASPGKPLDKLKORAAAI 974
      238 MTGSGEGEGEGEGEGSGDSFGVGAFGAPGSGQGERGRTGP-----AGQGRGA-- 289
      975 PLQVTKVHEPPREDAAPTKPAAPAPPPQNLQPSDAQDQPG--SSPRKSRSP----- 1026

```

```

      290 -----DGATSGQPPGSGTGP-----AGAPGPGISGAKGADGSPGARGSPG 330
      1027 -----APP-----ADKEAFAAAOQLPGDPPCMTGGLPVPVP 1059
      331 LOGARGRGSESGQGTGPPGVGGRDGSNGAKGSAPSSAQGTGPP--GARGPVP--- 385
      1060 REVITKSPHAPDDSATSVAPPCGHPRLGLHDTARPVLPPPTISNPPLISSAKHPSVLE 1119
      386 -----AGSPGPAGSKDQGNPGP--GAQESGFLGFGGTGPAPP----- 425
      1120 RQIGALSQGSVOLHVPVSEHAKAPVPGVTMGLPLPMWDPKLAFFSGVQEOQLSPGQAG 1179
      426 ---GA--QGSRGK--SRGALGPAGP-----GVGE--RGPMG 456
      1180 PPESLGVPTQAEASVLRGTLGSLVPPGSLTKGIPTSTRVPSDAITRGSLTHGTADVLX 1239
      457 PPGMSGAPGA-----PGAKGDRGLPG----- 477
      1240 KGTITRIGEDSPERLDRGREDSLPGKHVYEKKKGHVLSYEGGMSVTQCKEDRSSSG 1299
      478 -----ERGANGS--KSSAGESGRPG-----EPGM-----PGRGLTG 507
      1300 PPHETAAPKRTYDMMEGRVGRASSASIEGLMGRALIPPEHSPHHLKEQHHTGSTTQGI 1359
      508 PPGK-----QGRDGR-----PGPAGAPGPGNSGPGASGQGLGLV--GL 547
      1360 PRSYVEAOEVLREAEALILKREGTPPPPPSRDLTEAVYKTQALGPLKLAHGLVATVX 1419
      548 PGQOQ-----KEEGEDGQGSFGAPGLTGEGRKGEFVAGP--PGPGG----- 591
      1420 EAGRSIHIEPREELRHTPELPLARPLKEGSLTQGTPLKYDTGASTGSKKHVRSILGS 1479
      592 SAGRGNGGPGQ-----AGSMGPPGP-----PGASGDAQAQD--NGP 628
      1480 PGRTPPVNPLVDWADARALERACYE-----SLKSRPGTASSGGSSIAAGAVIYVE 1532
      629 PGESGPEGP-----GARG--ERGAPGERPGGLTGAGGRRKIGRAGNYGHTGPFGQKG 682
      1533 LGKPRQSLTYEDHGAAPAGHL--PRGSPVTREREPTPLQESLSSSSXASQDRKLTSTPRE 1591
      683 MGPPGNVGL--QGP--GELGSPGPPGANGPPGSGSPDPDPAGAEGR----- 728
      1592 IAKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTSLPRGILPDLAAAYVL 1651
      729 -----GVGP-----MGPSG-----PSGMGEEDG-- 748
      1653 PHLAPNTTYPHLYPPLIRGYPDTALENNQTIINDYITSQQMHNHTATMAQRADMLR 1711
      749 -----NBPFGG-----APQGRD-- 763
      1712 GLSPRESSLALNVAAGPRGIDLSQVPHLPVLVPPPTGTATAMDRLAYLFTAPQ---P 1767
      764 ---PD-----LGPQ-----SPSPFGAPGPRGSGNPGQSGELGP 796
      1768 FSSRHSSPPLSPGCTHLLTKPTTSSSERERDRERERERERERERERERERERERER 1827
      797 TGARGETGPGSPSGTDPGP-----OGPLG 822
      1828 RPTGSGSSGSSSGGCG-----GSSSRPASHAHQSPISPTODALQORPSVLANTG 1882
      823 APQOQGERGETPQOQGPGRPTGSLGAPGAG--PGPTGPSGMASSPGQPGARGBPG 879
      1883 MKGIITAVPSPKPTVLRST-----STSPVPRAATTPPATTCPLAGTLDGYVP 1930
      880 QSG-----SPQPGLAGRTGSGERGDKNDAQSPGPBPAGPAGQSGILGLAG-- 931
      1931 TLMERVLPLKKAAPRARPERPADGHAFLAKPPARSG-----LEPASSP----- 1975
      932 -----SGPR--GPGGPAAGPPAAGSRGAPGSGDRSGPAGVAGNPGRGENGM 979
      1976 --SKGSEPPPLVPVSGHATITARTPAKNLAPHHASPPDPAP-----PASADPHREKTQS 2028
      980 PGSDGNDGAP--GPGSRGKEKDDTASGA--NGSPGAPGFIGADGAGASGPR----- 1028

```


QY 545 LKERTDSTGSDNDEKEAVASGKRTANSQGRKRGRITRSMANSEATITPOQSAL 604
Db 547 QAEBSGKEG--SSEKEBGEQEGTEA-----EAGEEA-----580
QY 605 ASMEINSSRWMEEMETAKGLLEHGRNWSAIAWVGSKTVSOCKNFYNYKKONLDE 664
Db 581 --EAKKEKVEKEBEVATK-----BE 600
QY 665 ILQOHLKMEKERNARRKKKAPAAASEAPPPVEDEMEASGVSGNEEMVEAEAL 724
Db 601 LVAD--AKVEKBEKAKSPVPKSP-----VEEKGS-----628
QY 725 HASNGEVPREGESGATVNNSSDTESIPSPHTEAKOTQNGPKPATLIGADPPRPPT 784
Db 629 ----PVPK-----SPVEKKG-----SPVPSPVEEKSGSPVPS 659
QY 785 PPRTSRARIFPTPASEATGAPTPPAPPSAPPPV--VPKEKEETAAAPVE--E 839
Db 660 PVEBKGSIVSKSPVEBKAKSPVPSKPAVEAKSKAEVKGQKEBEKEVKAPKEKYE 719
QY 840 GEEQKPPAAELAVDTGAKEEVPKSECTEA-----870
Db 720 KKEEKPVPE---KKKASPVKEBAVAEVVITKSVKHLKETKEEKPLOQEKKE 775
QY 871 -----EEGPAG---KDAEAEKTAEG--AKAEKEGGSGRATTAKSSGAPD 914
Db 776 KAGGEGSEEGSDKGAKSGRKEDIAVNGEVEGKEBEQETEKESGR-----BE 825
QY 915 SDSATCSAD--EVDAAEGGDKN 935
Db 826 EKGVTNGDLSLPADEKKGDKS 848

RESULT 83
T32008
Hypothetical protein K106.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32008
R:Davidson, S.; Wohlmann, P.; Mullen, G.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of *C. elegans* cosmid K106.
A:Reference number: Z21111
A:Accession: T32008
A:Staccus: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1819 <DAV>
A:Cross-references: EMBL:AF016669; PIDs:ABE6098.1; GSPDB:GN00020; CESP:K106.3
A:Experimental source: strain Bristol NZ; clone K106
C:Genetics:
A:Gene: CESP:K106.3
A:Map position: 2
A:introns: 41/2; 191/2; 556/2; 701/2; 1645/2; 1700/2; 1768/1; 1797/3

Query Match 2.5%; Score 329.5; DB 2; Length 1819;
Best Local Similarity 19.2%; Pred. No. 0.00022;
Matches 366; Conservative 228; Mismatches 679; Indels 637; Gaps 87;
QY 815 PSAPPVVPEKEKEETAAAPVEEGE--EQKPPAAEELAVDTGAKEEVPKSECTEAEAG 873
Db 114 PQRPPPPQVDQSSSTGSLPPLHYRYQLPPRPBNQRTPTQOLQYVK--VVEA--GG 169
QY 874 PAKGDAEAALAVGALKAKEKSGSGRATTAKSSGAP-----QSDSATCSADEVD 927
Db 170 QAVAAQVQAQGS-----NRSGAAGVNSALQKPLPLPLSLTSSSSAAGSSISAPS 221
QY 928 EAEAGDKRLSPRBSLLPTGDP--ANAS-----PQKLDLKQKQRA 971
Db 222 TSQPTSSLSLTPSPSTSSSMAPRKTPPNASSSSSLIKQSQDVOEQQVDFEVANVSQ 281
QY 972 AIPPIQVTVHEPPPPREDAPTPKAPAPPPPNLQESDAPQOPGSSPR-----1020

Db 282 IWSKGLKVMHEPRLTGSLLP--QLAELAPLP-----PKSGVYQCPNCNRLANARNLQRR 336
QY 1021 ---GKSRRPAPADKAPAAEQKLPGRDPCPTSGLPFPVPRREVITKASPHAPDSAFY 1077
Db 337 QTCGAQHAAP---QLAAMLQRP--PCASA-----PP---VAPTPASTSFQHN 379
QY 1078 APPGHLPLGLHDTA-----RPVLPRPPTISNPPPLISA-----KHPVLE 1119
Db 380 NSTGN-LTISYSSSSSRHOSSLYSPOLEHODLVGNPNMLSGVEYKODPMLYQSPSGLS 438
QY 1120 ROIGALISQMSVOLHVPSSEAKAPVGVPTGLPLM--DP-KKLAPFSGKQOQLSPRQ 1177
Db 439 DSIV---SRDSSPHSPSPASHDQDMD--HLGFPDPLDPLHLHLSFDSADHKEPREK 494
QY 1178 AGPPESLGV-PTAQAASVLR--GTALGSVPGSITKGIPISTFVPSDASITTYGSLTHGP 1234
Db 495 HEPDELMTLDPTPPCGSERFYGINIDMP---LSLDCDEPLMRSEASLSSSQGRNP 551
QY 1235 ADVLYKGTITRLIGDSPRLDRGRDLSLPGHVIYEKGKGVLSYEGMSVYQCSKEDG 1294
Db 552 AAVFTRPPTTK-----562
QY 1295 RSSGPPHETAARKTYDMMEGRVGRASISIBGLMGRALPBRHSPHMLKEQHHR-- 1352
Db 563 -----RKPAKRPKSKKEASBEPKNSAILAAL--RKEPAAPQOPOLOFOQNYQPS 612
QY 1353 -----GSTGGIPRSTVEAQEDYLREALKLRBGTTPPPPSRDLTEYKIQALGP 1404
Db 613 POFQAPYGGSLPSTISASWLSHASTSAAAA-----PERSEMTSPIVTSAPNP 661
QY 1405 LKLKPAHEGLVATVKEAGRSIHEIPREELRHTPELP-----APRDLKEGS 1450
Db 662 Y-----IHQHPQOPQOKSSPLEBLNEQESADDDDSRSSST 702
QY 1451 ITQGTPLKYDTGASTGSKNDVRSLLSGPGTFPPVPHLDVMADARALERAQYSESLKS 1510
Db 703 VNSTT--TTTATTTSSKS-----TGNP-----LFTCEHCARQLCSMNLKR 743
QY 1511 RRGTT---ASSGGSITAR-----GAPVIVP-----ELGKPROSL--T 1542
Db 744 HRATKVAASSSNSNAASRPSPQSTPATATPMLQASQADQPLQAPQOSMETTAVT 803
QY 1543 Y-EDGAPFAGHLPRGSPVTMEPTPRLO-----EGSLSSS-----KASQDKLTSTPRRI 1592
Db 804 YTKTTPPSVANTWNTTEKAQLISPKPRQSTISESSSMTVVDALRAQHQKMQOQIOI 863
QY 1593 A-KSPHSTVPEHH-----PAPISPYEHLRGVSGVDLYRSHIPLAFDPTSPRGI 1641
Db 864 OFQOQOQOQFQHHQOQOQAGRIPPRPPIINQVNPQOVQHNQHMQLNPIR-----918
QY 1642 PLDAAAAYTLPRHLANPPTVPHLYRPPYLIRGYRDTALENRQTIINDVITSQOMHNTAT 1701
Db 919 -----QPLQSPPP-----PP-----PKGLLEHKT--DVLV-----TSE 948
QY 1702 AMAORADMLRGSPRESSIALNYAAGRGIIIDLSQVPHLVVPPPTGATAMRLAVL 1761
Db 949 PLAEKMD-----AKRRS-----EGLVAVTSTPLPIQLPQRSQAPARSQOQOQRP 994
QY 1762 PTAPO--PSSSRHSSPLSPGPTLTKPTT-----1791
Db 995 PVAQVQVQFNGR---PLPPMQLPPLQPNHQOQOQHMHQSQMNYQOVQOVQOVHQQO 1050
QY 1792 -----SSSERERDRER-----DRDRERKSLITSTTT-----VE-H 1823
Db 1051 QNLQNHQOQOQHOQOQOQAGNRSRSHSNVKNKEOEORQSGPLDSITISVPLSTEVH 1110
QY 1824 APIWRPGTEQSSGSGSGSGSGSSRPASHSAHQHSPISRQTDALQORPSV-----1877
Db 1111 HHIMKPRPLQGOSSVDOSGTABEPKRAAS---QAVICECKTASRAKNVGRHMA 1165
QY 1878 LHNTGKGIIT-----AVEP-----SKPTVLKSTSSSPVRPAATPPATHCPGLG- 1923
Db 1166 VHKLTLDLILANPEQPLDPLSAVQAGRRHTVAGLEPDSALTKAPATKRKASEAPSAQV 1225

```

QY 952 RANASPOKELDLKOLKORAAAIPIQVTKHNEPREDEADPTKAPAPPPQNIQPESDA 1011
Db 491 STTESSAPVTSTTESSAPV-----TPSSSTTESSA-PATPSSSTTESSA 540
QY 1012 P-----QOPGSPRGKSRSPAPPADKEAPAAEQKLGDDPCWTSGLPFPVP 1058
Db 541 PVTSTTESSAPVPTPSSSTTESSSTVTSSTTESSAPV-PTPSSSTTESSAPVPTP 599
QY 1059 PREVIK-ASPHADPSAFVAPPGHPLPLGLHDTAPVLPFP-----PTISNPPPLISSA 1112
Db 600 SSSSTTESSAPAPPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSS- 658
QY 1113 KHPSTLEHQIGALISGCMVQGLHVPYSEHAKAPVGVMTMGLPLPMDPKLAPFSGVKQQL 1172
Db 659 -----STTESSAPVPTPSSSTTESSAPVTS-----TTTESSAPVTSSTTSS 703
QY 1173 S-----PRQAGPRESLGVPTPAQEAASVLGTLGSPGSLTKGIPSTRVPSDAIYRG 1227
Db 704 SAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTE-----SSAPVTSSTTSS 760
QY 1228 SITGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVYISGKGHVLSYEGCMGV 1287
Db 761 SAPVPTPSS-----STT 772
QY 1288 QCKED-GRSSGPRHETAPKRTYDMMEGRVRAISSASIEGLMGRALPERRHSPHLK 1346
Db 773 ESSAPVPTPSSSTTESSAPVPT-----PSSSTTESSAPVPTPSSS----- 816
QY 1347 EQHNRGSIITOGIPRS--YVEAQEDYLRRKALKREGTPPPPSRDLTEAYKQALGP 1404
Db 817 -----NITSSAPSTPSSSTE-----SSSVPTPSSSTTES-----SSAP 853
QY 1405 LKLKRAHEGLVATYKEAGRSIHEIPREBLRHTPELPLAPRLKEGSIITGPKLYDTGAS 1464
Db 854 VS-SETTESSAPV-----PTPSSSNITSSAPSIIP-SS 887
QY 1465 TTGSKGHVRLIGSPGRTPFPVHLDVMAADARALERACYEBSLKSRRGTASSGGSIAR 1524
Db 888 TTES-----FSTGTTVP-----SSSKYRGSQETISVS-ST 917
QY 1525 GAPVIVP-----ELCKPROSLTYE--DHGAPFAGHLPRG--SPVMTREPTPLQEGSLS 1576
Db 918 TETITVPKTTSTVPTSTTTTITTCSTGTNSAETTSKCPKVTVTPTTTTSTT 977
QY 1577 SKASODRLTSTPREIAKSPHSTVPEHHPISPEYHLIRGVSGVDLYRSHIPLAFDPTS 1636
Db 978 SSTT--TITTT--VCSTGTNSAGE-----TTSKC-----SPT 1006
QY 1637 IPRGILPDAAAAYVLPRLAPRPYPLYPPLINGVDTALERQTIINDYITSQMH 1696
Db 1007 ITTVPCSTSPETASSTTSTSTTP-----VTTVSTTVVTTSTSTSTPG 1053
QY 1697 HNTATAMAQRADMLRGLSPRESSLALNYAAGRGIILDSOVPHLVLPPTGPATAMD 1756
Db 1054 GEITTFPTK-----NIPITY-----LTTIAPTSVTTVNTFTTTT 1092
QY 1757 RLAVIPLAPQPPSSHHSSPLSPGPPHLPKPTTSSSERERDRERDREREKSLT 1816
Db 1093 TVCSGTGN--SAGETTSKCPKVT--TTPVPCSTGTEVYTE-----ATTIVT 1136
QY 1817 S--TTTVEHAPIWRGTQSSG--SSGSSGGGSSSPASH-SHAHQHSPISTPTOL 1871
Db 1137 TAVVTV-----VTTSSSTGTNSAGKTTGYTTKSYPTTVTLADAPVTPATNAV- 1188
QY 1872 QORPEVLN-----TGKGIITAVEPSKPTVLRSTSTSSPVPRAPATEPPATHCLGSL 1925
Db 1189 ---PTTITTECSATNNAGETTSVCSAK-TIVSSASAGENAPATTPVTATP----- 1239
QY 1926 DGVVPTLMEVLLPREAPRVAPRPRADTGHAFIAPKPPAR-SGLEPSSPEKSGEP-RP 1983
Db 1240 -----TVITTTES--SVGTNSAGETTTGYTTSIPTTYITLILPGSNGAKNYEVAT 1289
QY 1984 LVPPVSGHAT--IART-PAKONLAPHHASDRPAPRANAS 2019

```

```

Db 1290 ATNPISIKTISQLATITASASVAPVPTSPSLTGLQSHS 1328
RESULT 82
A27864
Neurofilament triple M protein - human
N.Aternate names: NF-M (medium) protein
C.Species: Homo sapiens (man)
C.Date: 30-Jun-1989 #sequence revision 30-Jun-1989 #text change 13-Aug-1999
C.Accession: A27864; A30157
R.Myers, M.W.; Lazzarini, R.A.; Lee, V.M.Y.; Schlaepfer, W.W.; Nelson, D.L.
EMBO J. 6, 1617-1626, 1987
A.Title: The human mid-size neurofilament subunit: a repeated protein sequence and the r
A.Reference number: A27864; MUID:817275853; PMID:3608985
A.Accession: A27864
A.Molecule type: DNA
A.Residues: 1-916 <MYE>
A.Cross-references: GB:Y0067; NID:935045; PIDN:CA68276.1; PID:935046
R.Lee, V.M.Y.; Otvos Jr., L.; Garden, M.J.; Hollist, M.; Dietzschold, B.; Lazzarini, R.A
Proc. Natl. Acad. Sci. U.S.A. 85, 1998-2002, 1988
A.Title: Identification of the major multiphosphorylation site in mammalian neurofilament
A.Reference number: A30157; MUID:88158120; PMID:2450354
A.Contents: annotation; phosphorylation sites
C.Genetics:
A.Introns: 360/3; 402/2
C.Superfamily: cytoskeletal keratin
C.Keywords: coiled coil; phosphoprotein
F.615,628,641,654,680/Binding site: phosphate (Ser) (covalent) #status predicted
Query March 2.5% Score 329.5; DB 2; Length 916;
Best local similarity 19.24; Pred. No. 0.0001;
Matches 200; Conservative 149; Mismatches 361; Indels 333; Gaps 40;
QY 2 SGTOLVAQWTRATEPRYPHSLSPVOIARTHTDVGILEYQHSRDYASHLS----- 54
Db 30 SPSSGFRSQSMRSRSPSTVSSYKRSMLAPRLATSSAMLSAESLDPGSSSLNGSG 89
QY 55 PGIITQPRRRPRLSEFPQENERSQELHLRPESHSLVPELGKSEMEFIESKRRLLELP 114
Db 90 PGCGYKLSR-----SNEKEQLQGLNDRFAGYI--EKVHLEQNKKEIR-- 130
QY 115 DPLRSPPLATGPASESDITKD-----RSLGKLEPVSPSPHPDPLELVPRLS 168
Db 131 -----HEIQLRKQKASHAOLGDAYDOETHELATLBMVN-----HEKAQVQDSHL- 178
QY 169 KEELIOMDRVDREITWEEQOISKLKKQQLBEEBAKPRPEKVPSP-----IESKH 223
Db 179 EEDHLRKEEEREARLDITDEAIPALRKDIEEASLVKVELDKVQSLQOEVAFLRSNH 238
QY 224 RS-----LVQIITYENRKAEMARILBGLQPVLEPLYNOPSDTROHENIKINQM-- 276
Db 239 EEEVADLLAQI-----QASHIVE-----RKQYLTDISTALKE 272
QY 277 -RKKLILYFKRNHARKOMKQPCORYDQLEALEKVERIENPRRAKESKYREYEK 335
Db 273 IRSQSHESDNMEQAEEM--FKCRYAKLTAEBONKEAI-----RSAKI-ELAEIRKQ 323
QY 336 -----QPEIRKORELOERQWQSVGQSGLSMAARSEHVESEIIDGSEQEN----- 384
Db 324 LOSKSIELESRGKTESLERQLSIEE-----RHNDLSYQDTIQLENELRG 373
QY 385 ---LEKQMR-----QLAVIPLVADQORIFIMNGMLADPMKYVDRO 427
Db 374 KWEAARHIREODULLNVMAIDIEIAYRKL-----EGETRFSTFAGSITGLPVTTHRPPI 430
QY 428 VMNMMSDEKPTPRKPMQHPKNGFLASFLERKTVAE--CVLYUYTLKKVKNYSKLVPR 485
Db 431 TTSKIQTKTAEAPLKVQHK-----FYVEITIEETKVEDKSEMEBALTAITELAASKKE 486
QY 486 SYRRRGKQQQQQQQQQQQQQQQQQQPMPRSSQEKDEK-EKEKAKEKEKEKEPVENDKD 544
Db 487 EKEKAKEKEKEPEABEEVAKKSPLVATAPFYKEEGGRGEEBEGGEEBEEDEGAASD 546

```

RESULT 81
548478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)

N:Alternate names: extracellular glucosylase; mucin-like protein MUCL; protein YIR0199
C:Species: *Saccharomyces cerevisiae*
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S48478; A26877; B26877; S27281; U66123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: GB:247047; EMBL:Z38061; NID:G603997; PID:G763364; GSPDB:GN00009; MIT
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of SPAL.
A:Reference number: A91831; MUID:87194600; PMID:3106330
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:ML6164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YAZ>
A:Cross-references: EMBL:ML6165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Saccharomyces cerevisiae*
A:Reference number: S27281; MUID:89031230; PMID:3141213
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:G4551; PIDN:CAA32069.1; PID:g4552
R:Rambrichts, M.G.; Bauer, F.F.; Marmur, J.; Pictorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudohyphal growth of *Saccharomyces cerevisiae*
A:Reference number: U66123; MUID:96323237; PMID:8710886
A:Accession: U66123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <IAM>
A:Cross-references: GB:U00626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C:Genetic8
A:Gene: SGD:MUCL; STA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MIPS:YIR019C; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
E/5-21/Domain: transmembrane #status predicted <TM1>
E/1350-1366/Domain: transmembrane #status predicted <TM2>

```

Query Match Similarity      2.5%; Score 330; DB 1; Length 1367;
Beef Local Similarity     19.5%; Pred. No. 0.00015;
Matches      277; Conservative    161; Mismatches   601; Indels   380; Gaps    58;

QY      682 KKKAPAAASEAAPPVEDEMEAGVSGNEEHWEEALAHAGNEVPREGCGCAT 741
          ::::: : : : : : : : : : : : : : : : : : : : : : : : : :
DB      209 KSSTTSSSTSSSTTSSSTSSSTTSSSTSSSTSSSTSSSTSSST-----TATAT 261

QY      742 VNNSDTSISPHTEAKDTGONGPKP-----ATGADGPPEPGPT 784
          : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      262 PTTISCTKEKPPPTTS--CTKEKPTPHHDTPCTCKKTTSCTCKKTTTTPVTSS 319

QY      785 PERRTSRAPIEBTPAS---EATGAPTPPADPPSPGAPPVVFKEEKETAAAPVEEGE 841
          : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      320 STTESSSAPV-PTPSSSTTESSSAPVTSTTESSSAPVPT--PSGSTTESSSAPVTSST 376

QY      842 EOKPPAAELAVDTCKAEFPVKSECTEAEGSPAGKXDAABAA-----TAEGALKA 893
          : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      377 ESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTSSTTESSSAPVTSSTTESSSAPVTS 436

QY      894 EKKEGSGGRAT--TAKSGAPQDDSDSATCSADEVDAAEGGDKNLLSPRSLLTPTGDP 951
          : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      437 STTESSSAPVTSSTTESSSAPVPTPSSST-----TTESSSAPVTSSTTESSSAPVPTPSS 490

```



```

QY 804 GAPT-----PPAPPSBAPPPVVKKEKEBETAAPVEEGEOKPPAAELAVDQX 857
Db 216 GAPPQGGQNGNGEGBEPPVSGPMGR-----GPP-----GPPKKE-----DDE 256
QY 858 AEPVSECTEBAEBAKDAEAAEATAGALNAEKEGGSGRATTAKSGAPQDS 917
Db 257 AKRPKA-----GEGPFGPGQARGPFG-PGLPVKNGRGVPLGDAKGEAGAP----- 305
QY 918 SATCSADEVEAEGDKNLLSPRSLTTPGDPNANSPQKRLDKOLKORAAIPIQ 977
Db 306 -----GVKGE-----SSPGENGSP-GPMGRGLPGERGRTGAP 339
QY 978 VTKNHEPRREDAPTKRAPAPPPONLOPESDAQOQSS-----PRKSR 1024
Db 340 AAGAR---GNDQPPGAGPPGVGPPGPPGAPAKGEAPGTARGEGAQGGEG 396
QY 1025 SPAPPADKEAFAAEOKLPEDPCWTSGLFPVPREVIKASPHAPDPSAFSAPRPHL 1084
Db 397 TFGSG-----PAGASGNP-----TDGTP-----GAKSAGAPGLA-----GAPRPG 435
QY 1085 PLGLHDTAPVLPRPTTISNPPLISSAKHPSVLEROIGALISQMSVOLHVPYSEHAKAP 1144
Db 436 PRG-----PPD-----PQG-----ATGP 448
QY 1145 VGPV-TMGLPLPMDPKKLAPFSGVQKQOLSPGQAGPSSLVPTPAQELSVLRGTAAGV 1203
Db 449 LBPQKQOTGR-----GIAGFKGEQ-GPKGEPPGAPGAPGAPAGEEGKRG-ARGE 496
QY 1204 PGCSITTKGIPSTR-VPSDAITYRGIT-HGTPADVLYKGTITRLIGEDSSRL----- 1255
Db 497 PGGVPIGPBERGAPGNRGPFGQDLAPKAP-----GERGSGLAGAPK 543
QY 1256 ---DGR-EDSLPKGHVIEGKKGHVLSYEGMSVTOCKEDGSSSPRHETAA----- 1306
Db 544 ANGDRPGEPLPGARGL-TGRPDA-GPQGVGSGAPGEDGR--PBPPOGARGOP 599
QY 1307 ---PKTYDMEGRVRA-----ISSASIEGLMR-----AIPERISPHIKQ 1348
Db 600 GWMGPPGPKA-----NGEPKAGEKGLPGAPGLRGLPGKDGTEGAPGAPAGEE 655
QY 1349 HHIRG-SITOGIPRSVYAEQVLYREALKLKREGEPPP-----PPSRLT-EAVYTOA 1401
Db 656 QGAPPSGQGLP-----GPPGPPGSGKRGDQGVGEAGAPL 694
QY 1402 LGPL-----KLKPAHEGLVATYKAGRSIHEIPRELRHTPELPLAPRLKEGSIQ 1453
Db 695 VGPBERGPPGERGSPGAQGL-----QGRGLPRTGTGPKASGAPGAPGAPGQ 749
QY 1454 GTPPLKDTGASTTGSKKD-----VASLIGSPR-----TPPVHPLDVMADARLEA 1502
Db 750 GMP--GERGAGIAGPKDGRGVGEGPPGAPKDGGRGLTGTIGPPAGANG----- 801
QY 1503 CYESLKSPPGTAASSGSGSIARGAPVYELGKPROSLTYEDHGAPFAGHLPRGSPVTM 1562
Db 802 --EKGEVGPAGSAG--ARCAPREBETGPPGSGIA-----GPPGAD-----QGPAGK 848
QY 1563 REPPLQEGSLSSSKASQDRKLTSTREIANSPHSTVEHHPHISPYEHLRGVSDV 1622
Db 849 GEGEGEGQKGD-----AGAPGQGSAGAPGCP-----TGVTGPK 884
QY 1623 LVRSHPLAFDPTSTIRGILPLDAAAAYVLRHLAPRTYPLHLYPLTIGYPTALENR 1682
Db 885 GARG--AQGP--PGATGPPAAGRVPBGSGNGPBG--PP--GPPGSGKDP 928
QY 1683 OTINDYITSQOMHNHTATAMAORADMLKGLSPRESSALINYAAG----- 1728
Db 929 K-----GARGP--SGPPRAEBEPLQGPAGPGRGKGERGDDGPG 966
QY 1729 -----RGIIDL-----SOVPHL--PVLVPTPTGPTATAMDLAVLPTAPOP 1767
Db 967 AEGPPGQGLAGRGIVGLPGQGRGGRGFPGLPESGEPQOQAGFASGDR----- 1016

```

```

QY 1768 FSSRRSSPLSPGGETHLTKPPTTSSSERERDRDRERERESKILTSTVHEABIW 1827
Db 1017 -----GPPGVPGRGLT-----GPPG-----GPPG 1032
QY 1828 RPTGQSSGSSGSSGSSGSSSRPASHAHQHSFISPTODALQOPSVLHNTMKGI 1887
Db 1033 EPGREGSPGADGPPGRDGA-----GVKQ-- 1056
QY 1888 TAVESKTVLAKSTSTSPVPAATFPATHCPIG-----GTLDDVPLMBPV 1936
Db 1057 -----DRGETGVAGAPGAPPGSPGAPGTGKQDRGAGAGPVGPS----- 1100
QY 1937 LLPKAPRVAPERRADTGH-----FLAKPPASGLEPASPSPKG 1978
Db 1101 -GPAGARGLQGPGRGKGEGERGKLGHRFTLOGLPGRPGSGGQAGSAPRAG 1159
QY 1979 SEPRPLVPVSGHATIAITPAKNLAPHNASDPPAPASADPHREKTQSKPFSIOEL 2038
Db 1160 SGPR-----GPPGPVGPSS----- 1172
QY 2039 RSLGTHGSSYSPGEVPEVPSPSLTGDKLPHLELDKSHLEGELRPKQGPVXLGG 2098
Db 1173 ---GKDGANGIPLPGLPGRGPRGSGETGPAGPGR-----PBPGRPGP 1215
QY 2099 EAAHLPHLRPLPESQSSPPL-----QTAGVKGHQHV-----TLAQHISEVITODYT 2148
Db 1216 PGIDMSAPAGLGPGRKGDPLQVYMDADAGLRHDAEVDATLKSINNQIESIPSGS 1275
QY 2149 RHHPQOLAPLAPLPSFPGASCPVLDLRP--PSDLYLPPPDHAPARGSP--HSRGG 2203
Db 1276 RKNPAR-----TCRDLKLCHPEWKSQGYWIDPNOCTIDAMKVCNMBTG 1320
QY 2204 KSPREPNTSV-----LGG-----GEDEIEVSPREGTEGH 2236
Db 1321 ETCVYENPAPVKNKMMSSKSEKHIWFGETINGFHFSTGDDMLAPYNAVQIT----- 1376
QY 2237 SRSAVPLLYRDEQTEPSRMGSKSPGNTSOPPAFSLKTESNAMYKSKQKQINKLNT 2296
Db 1377 ---FLRLSTEGSN-----ITHCKNSIAYLDEAAGNLKALLI 1413
QY 2297 HNRNPEYNISQPTGTFINMPAIT-----GTGLMTYSQ 2330
Db 1414 QGSNDV--IRAEGNSRFTYTLAKDCTKTKGKMTVIEYRSQ 1455

```

RESULT 80
 CGHUV
 collagen alpha 1(V) chain precursor - human
 N.Alternate names: procollagen alpha 1(V) chain
 C.Species: Homo sapiens (man)
 C.Date: 22-Nov-1993 #sequence revision 03-Oct-1995 #text change 16-Jun-2000
 C.Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665
 R.Greenspan, D.S.; Cheng, W.; Hoffman, G.G.
 J. Biol. Chem. 266, 24727-24733, 1991
 A.Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of e
 A.Reference number: S18802; MUID:92105142; PMID:1722213
 A.Accession: S18802
 A.Molecule type: mRNA
 A.Residues: 1-1838 <GRE>
 A.Cross-references: GB:M676729; NID:G189519; PIDN:AAA5993.1; PID:G189520
 J.Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaol, Y.; Kato, I.
 J. Biol. Chem. 266, 13124-13129, 1991
 A>Title: Complete primary structure of human collagen alpha-1(V) chain.
 A.Reference number: S16024; MUID:91502336; PMID:2071595
 A.Accession: S16024
 A.Molecule type: mRNA
 A.Residues: 1-81, 'OU', 84-389, 'A', 391-676, 'K', 678-1294, 'PS', 1297, 'RS', 1300-1553, 'R', 1555-
 A.Cross-references: GB:D90279; NID:G219509; PIDN:BA14323.1; PID:G219510
 A>Note: parts of this sequence were determined by protein sequencing
 R.Yaol, Y.; Hashimoto, K.; Takahara, K.; Kato, I.
 Exp. Cell Res. 194, 180-185, 1991
 A>Title: Insulin binds to type V collagen with retention of mitogenic activity.
 A.Reference number: A61142; MUID:91224163; PMID:1709100

Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
 A>Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pro
 A:Reference number: A04227; MUID:90370826; PMID:1975693
 A:Accession: A33116
 A:Molecule type: DNA
 A:Residues: 171-172, 'C', 174-175 <ATA>
 A>Note: mutant sequence from a family with family with primary generalized osteoarthritis
 R:Diab, M.; Wu, J.J.; Eyre, D.R.
 Biochem. J. 314, 327-332, 1996
 A>Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
 A:Reference number: S64673; MUID:96195147; PMID:860302
 A:Accession: S64674
 A:Molecule type: protein
 A:Residues: 188-189, 'X', 191-195, 1224-1230, 'X', 1232-1236 <DIA>
 R:Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
 Eur. J. Biochem. 234, 125-131, 1995
 A>Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil car
 A:Reference number: S63514; MUID:96096730; PMID:8529631
 A:Accession: S63514
 A:Molecule type: protein
 A:Residues: 243-261, 575-590, 756-763, 'X', 765-779 <FRA>
 R:Tiller, G.E.; Weis, M.A.; Colombo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
 Am. J. Hum. Genet. 56, 388-395, 1995
 A>Title: An RNA-splicing mutation (G+51VS20) in the type II collagen gene (COL2A1) in a
 A:Reference number: 138867; MUID:95150028; PMID:7847372
 A:Accession: 138867
 A:Structure: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509 <TIL1>
 A:Cross-references: EMBL:X15195; NID:9557053; PIDN:AAB60370.1; PID:9557054
 R:Ramirez, F.
 submitted to the EMBL Data Library, December 1988
 A:Reference number: S04892
 A:Accession: S04892
 A:Molecule type: mRNA
 A:Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <RAM>
 A:Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050
 R:Viikula, M.; Peltonen, L.
 FEBS Lett. 250, 171-174, 1989
 A>Title: Structural analyses of the polymorphic area in type II collagen gene.
 A:Reference number: S05000; MUID:89325561; PMID:2753125
 A:Accession: S05000
 A:Molecule type: DNA
 A:Residues: 630-640, 'A', 642-785 <VIK2>
 A:Cross-references: EMBL:X16158; NID:929951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
 PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
 R:Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
 J. Biol. Chem. 267, 22522-22526, 1992
 A>Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(I) chain pro
 A:Reference number: A44309; MUID:93054548; PMID:1428602
 A:Accession: A44309
 A>Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA; mRNA
 A:Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T',
 A:Cross-references: GB:L00977; NID:9180812; PIDN:AAB3914.1; PID:9258774
 A>Note: sequence extracted from NCBI backbone (NCBI:117273); parts of this sequence wer
 A>Note: this translation is not annotated and this publication is not cited in Genbank
 R:Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
 A>Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individua
 A:Reference number: S16502; MUID:90251662; PMID:2333128
 A:Accession: S16502
 A:Molecule type: DNA
 A:Residues: 1164-1184, 'GPGKGDANGIPGP', 1185-1199 <TIL2>
 A:Cross-references: EMBL:M37126; NID:9180808; PIDN:AAA52037.1; PID:g180809
 R:Chenah, K.S.B.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
 Proc. Natl. Acad. Sci. U.S.A. 87, 2555-2559, 1985
 A>Title: Identification and characterization of the human type II collagen gene (COL2A1)
 A:Reference number: A02858; MUID:85190534; PMID:3857598
 A:Accession: A02858
 A:Molecule type: DNA

A:Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487 <CHE>
 A:Cross-references: GB:J00116; NID:g180395; PIDN:AAA5197.1; PID:g180396
 R:Elma, K.; Viorio, T.; Viorio, F.
 Nucleic Acids Res. 15, 9499-9504, 1987
 A>Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
 A:Reference number: A27280; MUID:88067771; PMID:2825137
 A:Accession: A27280
 A:Molecule type: DNA; mRNA
 A:Residues: 1175-1487 <ELI>
 A:Cross-references: EMBL:X05268; NID:930096; PIDN:CAA29604.1; PID:930097
 A:Experimental source: fetal epiphyseal cartilage
 R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
 Biochem. J. 237, 923-925, 1986
 A>Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
 A:Reference number: A57033; MUID:87099927; PMID:3800925
 A:Accession: A57033
 A:Molecule type: protein
 A:Residues: 'X', 1244-1246, 'N', 1248, 'X', 1250-1265, 1295-1305, 1395-1408 <VAN>
 A>Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prop
 R:Strom, C.M.; Upholt, W.B.
 Nucleic Acids Res. 12, 1025-1038, 1984
 A>Title: Isolation and characterization of genomic clones corresponding to the human typ
 A:Reference number: A21733; MUID:8418798; PMID:6320112
 A:Accession: A21733
 A:Molecule type: DNA
 A:Residues: 1245-1295 <STR1>
 A:Cross-references: EMBL:X00339; EMBL:X00298; NID:9394699; PIDN:CAA25092.1; PID:g4378975
 A:Accession: B21733
 A:Molecule type: DNA
 A:Residues: 894-909, 'PE' <STR2>
 A:Cross-references: GB:X01785; NID:930035; PIDN:CAA25082.1; PID:g1335032
 R:Nunez, A.M.; Francopano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
 Biochemistry 24, 6343-6348, 1985
 A>Title: Isolation and partial characterization of genomic clones coding for a human pro
 gene.
 A:Reference number: A24561; MUID:86104139; PMID:3002437
 A:Accession: A24561
 A:Molecule type: DNA
 A:Residues: 1296-1358 <NUN2>
 A:Cross-references: GB:M12048; NID:g180017
 A>Note: this translation is not annotated in Genbank entry HMWCT2A, release 111.0
 A>Note: the codons given for 1333-Lys (AAG) and 1350-Gly (GCA) are inconsistent with the
 R:Sangiorji, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
 Nucleic Acids Res. 13, 2207-2225, 1985
 A>Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll.
 A:Reference number: 137249; MUID:85215609; PMID:2987845
 A:Accession: S59491
 A:Molecule type: DNA
 A:Residues: 7-28, 'R', 99-114, 541-578, 786-802, 1055-1056, 'N', 1058-1068, 'T', 1070-1109, 1200-1
 A:Accession: 184453
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 7-28 <SAN2>
 A:Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
 A>Note: the Genbank PID is based on an incorrect reading frame
 A:Accession: 137250
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 541-560 <SAN3>
 A:Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
 A:Accession: 137251

Query Match 2.5%; Score 333.5; DB 1; Length 1487;
 Best Local Similarity 20.4%; Pred. No. 0.00013;
 Matches 377; Conservative 116; Mismatches 628; Indels 723; Gaps 85;

Qy	726	ASGNEVPRGEGSGATVNNSSDTSIPSPHTEAK-----DTGQNG-KPK-----P	770
Db	96	ASGQPGPGGQGGEDIDIVGPKGPGPGPAGGQGRGGRGKGAGAPGRGDSF	155
Qy	771	ATLGADGP--PPGPTPP-----BRTSRAPIEPTPASEAT	803
Db	156	GTLCNPGPGPGPPGPGGLGCPAQAQAGGFGDEKAGAGOLGWVQGPWGPGGPGPGPPA	215

```

Db      408 -----EEDGVRCGAPLARAGPPSP-----APAAP---RP 437
QY      1410 AHEGLVATYKEAGRSIHETPREELRHTPELPLAPRLKSGSTTGGTPLYDPGASTGSK 1469
Db      438 SASASANTSSSAASAPAPPE-----PARPPKRRKSTNNHLSLMADPPPTDG--- 486
QY      1470 KHDVSLTSGPRTPEPPVPLDVMDARALERACYEESLKSRRG-----TASSGGS 1521
Db      487 -----PLTLPGEPPMGSDP---PADGRVR-----YGAADSDREGLMDEDVQMAARVYA 534
QY      1522 IARGAPVIVPELCKPRQS-----PLTYEDHGAFFAGHLPRGSPVTNREPTPLQEGSLSS 1576
Db      535 AAGPVVVFIFEMWGDSCRQHEALVRLIYS--GA-----AGEAMSWLQ--NPRMQ----- 578
QY      1577 SAASODRKLTSTPRELAKSPHS-----TYPEHH-----PHPISP--- 1610
Db      579 ---APQRFNQCQRRVVAHPHGSGFITSGVTPPEPLHIGDMAAODPLMALPNAVASAVAM 635
QY      1611 ---YEHILRGVGVDTYRSHIPLAF-----DP-----TSIPRGPPL 1643
Db      636 SRRYDRTOQTPILOSIRAYADMAVYGRADDPAGAGATVEALCARVAAFAAAQPGRVVR 695
QY      1644 DAAAAAYTL-----PRLHAPNTPYPHLYPPYLLRGYPDTALENROTIINDYIT 1691
Db      696 ELADACVLACRGVLERLLPCPLRLPAPARAPALGPACLEEV--TAAI---LALRDALP 749
QY      1692 S-----QOMHNTATMAQRADMLR---GLSPRESSLALNYAA---GPRGID--LSQ 1736
Db      750 GAGPAPERQOADSVALVARTVAPLVYSDGARAARA--AMTYAALFAPAVNAGARLAE 807
QY      1737 VEHLPVLVPTPGTPTAMDRLAYLPTAPQPFSSRRSSPLSPGCGTHLTKPTTSSSR 1796
Db      808 AAARGPAPAPAGLPPLPMEQEGLVVPAAP--AAAGPBGGLGSGG---SSPASTSG-- 861
QY      1797 EEDRDRERDREREKSIITSTTVENHAPIMRPGT-----EQSSG 1836
Db      862 -----SSTKSSSGTSGSGSGVYARLPRRRPGSARAQAQBAPARACPDGDEBDG 915
QY      1837 SSGSSGGGGSSSRPASHAHOHSPISPTODALOQRSVLHNTMKIITIAVESKPT 1896
Db      916 LSGSALARGDG-----HGRDDEEDRGR---RKRSI---GLGPDPAPAL 956
QY      1897 VLKSTSS-----PVRPATFP-----PA--THC--PLGGLTDGVYPTIMBP 1935
Db      957 VSSSSSSSEDDRLRLPLGPMENHAPDGGRRVAGETHPRPAAALAAVCP----- 1011
QY      1936 VLLPKAPVAR-----PR--RPADDTGNAFLAKPPASGLBPASSPSKSGEBRPLV 1985
Db      1012 -----PEVARALVDQEVFPELWRPALTFDPALALHIAARCAAGAPLRRRAAMWRQIA 1064
QY      1986 PIVS-----GHATITARTPAKTLAPHNASPPRPAFA--SASPPHEKT----- 2026
Db      1065 DEDVAVVVLVLDPLPHEELCAPEAEQ--APRGGLPAPRGSLALLAFAHLCPTDISHAMAG 1123
QY      2027 -----QSKPFSIQELESLSLYHSSSYSPGEGEYPS-----P 2058
Db      1124 NWTGPRDICKLAAGVLLISARDLSAGA-----VEYICSLGAARRRLIVLDTIEDMP 1177
QY      2059 VSSPSLTMDKGLPKHLELDKSHLBEGLRPKQPGVYKLGAEAAHLPHLRPESQSSSP 2118
Db      1178 ADGPVAVG-----DYHV-----YVRARLDPAQCAVR-----WPECSELNAAVLDSSS 1219
QY      2119 LIQTA-----FGVKGHQVVTLAGHISVITQDTRHHPOOLS---APLPAP 2162
Db      1220 IYGPACFARVEASFARLHPGAP-----LTLICQDDVVRYVTSFRAAPRTVPVPLP 1268
QY      2163 LVSFPGASCPVL---DLRRPSDYLTPRPDGAAPRAGSPHSGGKRSPEPKNTSVLGGG 2218
Db      1269 PPAIYRERVLPYDGCCKDMARKORSALGDPDPGAGAFGRANRWGLGAPLRPVVVS CG 1328
QY      2219 EDGIEIVSPPEGTEPQSHRSAYPLLYDGEQTEPSSRGSKSPGNTSQPPA 2270
Db      1329 RRGIAELRSPREGI--PAELRAFCAALALBPDAEAAVLVLVAPGALMAAGAPFA 1378

```

RESULT 79

CGHUC

collagen alpha 1(II) chain precursor [validated] - human

N/alternate names: procollagen alpha 1(II) chain

M/contains: chondrocalcin, collagen alpha 1(II) chain precursor splice form 1; collagen

C/species: Homo sapiens (man)

C/date: 26-May-1986 #sequence revision 01-Sep-1995 #text change 08-Dec-2000

C/accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63

7250; I37251; I37252; I37253; I37254; I5538; I59535; I61910

R/Ryan, M.C.; Sieraski, M.; Sandell, L.J.

Genomics 8, 41-48, 1990

A/title: The human type II procollagen gene: identification of an additional protein-cod

A/reference number: A38513; PMID:91184811; PMID:2081599

A/accession: A38513

A/molecule type: DNA

A/residues: 1-103 <RYA>

A/cross-references: GB:M60299; NID:g180883; PIDN:AAA73873.1; PID:g180884

R/Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.

Nucleic Acids Res. 17, 9473, 1989

A/title: Nucleotide sequence of the full length cDNA encoding for human type II procolla

A/reference number: S06715; PMID:90067946; PMID:2587267

A/accession: S06715

A/molecule type: mRNA

A/residues: 1-28, 'R', 99-1487 <SU2>

A/cross-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID:g29516

A/Note: alternative splice form 1

R/Viikula, J.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.

Biochem. J. 285, 287-294, 1992

A/title: Structural analysis of the regulatory elements of the type-II procollagen gene.

A/reference number: S24270; PMID:9244585; PMID:1637314

A/accession: S24270

A/status: translation not shown

A/molecule type: DNA

A/residues: 1-28 <VIK>

A/cross-references: EMBL:X58709; GB:S40537; NID:935659

A/Note: this translation is not annotated in Genbank entry HSPROCOE1, release 111.0

R/Nunz, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.

Gene 44, 11-16, 1986

A/title: Promoter region of the human pro-alpha-1(II)-collagen gene.

A/reference number: A24828; PMID:87031574; PMID:3021582

A/accession: A24828

A/molecule type: DNA

A/residues: 1-8, 'T', 10-28 <NUN>

A/cross-references: GB:M25698; NID:g180872; PIDN:AAA52051.1; PID:g553237

R/Balwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.

Biochem. J. 262, 521-528, 1989

A/title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)

A/reference number: S06496; PMID:90026318; PMID:2803268

A/accession: S06496

A/molecule type: mRNA

A/residues: 7-28, 'R', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'F'

A/cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041

A/Note: alternative splice form 1

R/Ryan, M.C.; Sandell, L.J.

J. Biol. Chem. 265, 10334-10339, 1990

A/title: Differential expression of a cysteine-rich domain in the amino-terminal propept

A/reference number: A35428; PMID:90285153; PMID:2355003

A/accession: A35428

A/status: not compared with conceptual translation

A/molecule type: mRNA

A/residues: 27-81, 'L', 83-103 <RYA2>

A/Note: alternative splice form 2; splicing appears to be under developmental regulation

R/Su, M.W.; Benson-Chanda, V.; Visseling, H.; Ramirez, F.

Genomics 4, 438-441, 1989

A/title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf

A/reference number: A30147; PMID:89233138; PMID:2714801

A/accession: A30147

A/molecule type: DNA

A/residues: 104-157, 'P', 159-236 <SUN>

A/cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GB

R/Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.

Db 223 AGAP-----GVKG-----SGSGENGSP-GRMGPRGLPG 261

Qy 969 RAAALPIQVTKVHEPPREDAAPTKAPAPPPQNLQPSDAPQOPSS----- 1018

Db 262 ERGRGRPGAGAACAR---GNDQOPGAGPPGVPGRGAGGPPGAPABAKKEAGPTGARPEG 318

Qy 1019 ---PRGKSRSPAPADKEAFAAEAKLPPDPECWTSGLPFVPPREVIKASHAPDPSAF 1075

Db 319 AOGPRGEPTGPGSG-----PAGAGNPG-----TDGIP-----GAKSGAGAGIA- 359

Qy 1076 SYAPRGHPLUGLHDTAPVLPPTISNPPLISSAKHPSLEROIGALISQMSVOLHU 1135

Db 360 --GAGFGGPRG----- 369

Qy 1136 PYSEHAKKPVGVNTWGLPLPMDPKKLA---PESGVKOBOLSPRGQAGPESLIGVPTAOEA 1192

Db 370 -----PPGD--QGATGGLGPPKQOTGERPIAGKKGEO-GRKSGPRGAGPGAGPPAGE 418

Qy 1193 SVLRGTALGSVPGSGITKGKIPSTR--VPSDSAITYRGSIT--HGTPADVLYKGTITRIIGE 1249

Db 419 EGKRG--ANGE--PGGAPVGPPEGRGAPGNRGPPGDGLAGPKAP-----GE 463

Qy 1250 DSPSL-----DRGR--EDSLPKGHVITYEKKAGHVLSTEGMSVTQCSKEDRSSG 1299

Db 464 RGPSTLAPRKGANCDPRGPRGEBGLPGARGL--TGRPGDA--GPGSKGPGSAGPEDGR--PG 519

Qy 1300 PPHETAA-----PKRTYDMMEGRVGA-----ISSASISGLMR-----AIP 1337

Db 520 PPGPGAGAQPGVWMPGPGPKA---NGBRPGAEGKGLPGAGLGLPKQDGETAGADP 575

Qy 1338 ERHSPHHLKEQHNRIG--SITQGIPRSVEAOEDYLREAKLKREBTPPP-----PESR 1391

Db 576 GPAGAGAGEGEGAGPGSGFGLP-----GPGPRPEGGKPGDQ 614

Qy 1392 DIT-EAYKTQALGPL-----KLKPAHEGLVATVKEAGRSIHNI PREELNHPLEPLA 1442

Db 615 GVPGEAGAGLGLVPRGERGFPGERGSPGAQGL-----QGARLPPTGTDGPKGASGAG 669

Qy 1443 PRPLEGASTTQTPPLKYDTGASTTSGSKND-----VRLSLGPR-----TFRVHPND 1491

Db 670 PPGAGGPPRLQOMP--GERGAGAGIAGPKDRDVGEBKGEPAKDXGKGGLGPIGPP 727

Qy 1492 VVADRALERACYESISIKSRPTASSGSGSIRGAPVILPELGKROSLTYEDHGADPA 1551

Db 728 AGANG-----EKGEVPPRPAGTAG---ARKAPGERGETGPP--GP-----AGRA 767

Qy 1552 GHLPGSGPVTKMEPTPRLOEGSLSSKASQDRKLTSTPREIAKSHSTVPEHHHPISPY 1611

Db 768 G--PGAD--GQPGAKGEG-----EAQ-----KDGAGAPGQPGSAGPQGP- 808

Qy 1612 EHLIAGVSGVDLYRSHIPLAPDPTSPRGIRPLDAAAYLPHLAPNPTPHLYPYLIR 1671

Db 809 ---TGVTGPRKARG---AGGP---PGATGPGGAAGVPGSGNGNP-----PP--- 848

Qy 1672 GYPTAALENROTINDYITTSOOMHNHTATAMAQRADMRLGSPRESSLALUYAAGP--- 1728

Db 849 GPPGSGKDGPK-----GARGD--SGPRGAAGDGLGLOGRPGE 886

Qy 1729 -----RGIDL-----SOVPHL--PVLVPTGTATAMD 1756

Db 887 KGEPGDGPSPGDPGPPGQGLAGORGI VGLPCORGERGPPGLPSPSGEPGAGAGASGD 946

Qy 1757 RLALYLPATQPPSSSRSSSPLSPGCTHTLTKTTTTSSRENRDRERNDREKRSILIT 1816

Db 947 R-----GPPGVGPPGLTG---AGSPRGEGTPGAGPGPRGRGAGV 985

Qy 1817 STTVHEAIRMPTGTEOSSGSSGSSGSGSSR-----PASHNH-ONSPLSP 1865

Db 986 KDDRGEAGLAPGAPGPPGSPGAPRTKQODRGAQOGVMGRAPGAGALGPGOP 1045

Qy 1866 RTQDA---LQORPSVLYN--TGMKGITAVBSKPTVLRSTSTSSPVPRPA--TEPPATH 1918

```

Db      1046  KGDGEGAGEACERLKLKHGFTGIQGL-----PDPSPSDQAGSAGBGPSPGPPGPV 1100
QY      1919  CPLGGITLDGVYPTLMEVLLPEKAPVAREPERRADTGHAFLAKPARSGLEPASPSKG 1978
Db      1101  GPFG--XDA-----NGIPGPIGPFGPRGSGETGPAPFGNPG----- 1137
QY      1979  SEPPPLVPVPSGHTTARTPAKNLAPHNASDP-----PAPPASDPHEKTOSKEFSI 2033
Db      1138  -PPSPGPPGPG--IMSAFVAGLGPKEKGPDPLOYNRADEAAGLRPHDEEV----- 1186
QY      2034  QEELRSISGYHSSSY-SPG 2052
Db      1187  -EATLKLNNQIESIRSPBG 1205

RESULT 78
EDBEIP
: immediate-early protein IE180 - suid herpesvirus 1 (strain Indiana-Funkhauser)
C:Species: suid herpesvirus 1
C:Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 16-Feb-1997
C:Accession: S04713
R:Cheung, A.K.
Nucleic Acids Res. 17, 4637-4646, 1989
A:Title: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies v
A:Reference number: S04713; MUID:89315207; PMID:2546124
A:Accession: S04713
A:Molecule type: DNA
A:Residues: 1-1460 <CHE>
C:Superfamily: herpesvirus immediate-early protein IE175
C:Keywords: DNA binding; early protein; transcription regulation

Query Match      2.5%; Score 335; DB 1; Length 1460;
Beet Local Similarity 20.4%; Pred. No. 0.0011;
Matches 341; Conservative 140; Mismatches 643; Indels 548; Gaps 71;

QY      832  AARPVEGEGEQKPPABELAVDTGKAEPFKSECTBAEGRA-----KKQDAAE-A 885
Db      22  AAAAEEEGGISGP-----DGSQSGRRSGSGEDLFGPGGLPSDDAAEEAAVLA 73
QY      886  TAEGALKAKKEGSGRATTAKSSG---APQSDSMTCSADVEDAEGDKNLLSPRP 942
Db      74  AAGATRRPPPSAQOQHNAKSGEIVLDDDE---EEDFSGSPAAGSPVGLSTRAP 129
QY      943  SLTPTGDPANASPOKEPLDLKOLKORAAIPIQ--VTKNPERRPEDIAETKEPARAP 1000
Db      130  STVSSSGPFGGAPRGPRRRQHSQRRPRPARAGARPRPRPRPARAPARAP 189
QY      1001  PPQLQPSDAPQPGS---SPRGKSRPARPADKEAFAEAQKLPQDPFCWTGGLPPPV 1057
Db      190  APR--RPRGDPGPRGSTRVSP--GRRRGLGRRIHQHSQQRVORHGGSP-----LPQRP 241
QY      1058  PPRVYKASPHAPRPSASVAPRGHRLYLGHDTAPRLPRPPTISNPRPLISSAKPSV 1117
Db      242  PP-----PGRSRPAAAPRAPAGTAVVTTSTASPLWDEPAAARLIDPAAARPREPL 295
QY      1118  LERQIGAISQMSVOLHVPYSENAKAPVGTWGLPLPMDCKLAPFSGVQEQQLSPRGQ 1177
Db      296  LQPG-----LQDLHNRRRRARPR-----RPRG-----RGR 323
QY      1178  AGPPSLGVP-----TAQDASVLRGTALGSPVSGSITGIGPSTVPSSAITTGRSI 1229
Db      324  TRPRGRGAPLQORPPRRRAGEGALRRGRGSSSSSGS----- 361
QY      1230  THGTADVLYGTTIRIIGEDSPRLDGRGDSLPKGVHVEGKKHVLSYEGGMSVTQC 1289
Db      362  -----SDSDLSPAKSPSPR----- 376
QY      1290  SKEDRSSSGCPHETAAFKRTYDMMEGRVGAISASISIGLMGRALPPRRHSPHLLKQH 1349
Db      377  -----AAAAAAARRGASSSSSSSSSSSSSSSEG----- 407
QY      1350  HIRGSIIGIPRSYVAQEDYLRBAKLKKEGTPPPPPSPSRDLTEAVKTQALPRLKPK 1409

```

Matches 330; Conservative 117; Mismatches 613; Indels 474; Gaps 75;

QY 725 HASGNEVRGECSPATVNS--SDTE-----SIPSPHTEAKDTGONGPKRPA 771
 Db 35 HAS--STGGEKETATQSSVSPSTEKNAVSMTSVLSHSPGSGSTTGQODVTLAPA 92
 QY 772 TLGADG-----PPPGPTPPRR--TSAPLIEPTPASA-----TGAPT 807
 Db 93 TEPAGSAATWGQDVTSVPTRPALGSTTPPAHDVTSAPDNKPAAGSTAPPAHGVTSAPD 152
 QY 808 PPPAPSPSAPPVVPKEKEEETAAAPVEEGEQKPAABELVDGKA---EEPKS 864
 Db 153 TRPAGGSTAPPAHGV---TSAPDTRPA---GSTAPPAHGVTSAPDTRPAAGSTAPPAH 205
 QY 865 ECTEEAEBSGAKGDAEAATAGALKAEKKEGSGRATTAKSSGAPODSDSATCSAD 924
 Db 206 GVTSAAPDTRPAAGSTAPPA---HGVTSAADTRPAAGSTAPPAHGVTSAPD 252
 QY 925 EVDABEGGDKRLSPRPSLLTPTGDDPRANASPOKPLDKOLKORAAIPIQ-VTKYHE 983
 Db 253 -----TRPAGSTAPPAHGVTSAPDTRPA---GSTAPPAHGV--- 288
 QY 984 PRREDAAPTTPAP-PAPPPQMLQESDAPQOGSSPRGKSPPAPADKEAFAPAAOKL 1042
 Db 289 ---SAPDTRPAAGSTAPPAHGVTSAPDTRPAAGST---APPAHGVTSAPDTRPA 336
 QY 1043 PGDPPCMTSGLPFPVPPREVIRASPHAPDPSAFYAPGHPPLPLGLMDTARFVLPPTI 1102
 Db 337 PGS-----TAPPAHGVTSAPDTR-PPAGSTAPPAHGV---TSAPDTRPAAGST 380
 QY 1103 SNRPPLISSAKHPSVLEROIQALISQMSVOLHVPSSEHAKAPVGPVWMLPLPMDPKLA 1162
 Db 381 APPPAHGVTSAP-----PDRTPAGSTAPPAHGVTSAPDTRP-A 416
 QY 1163 PPSGVKQEOQLSPRGQAPRESLGVPTAOBASVLRGTALGVSFGSITTKGIPSTR-VPSDS 1221
 Db 417 P-----GSTAP-APGVTSAPDTRPAAGST-APPAHGVTSAP-PTTRPAAGST 460
 QY 1222 AITTRGSIHGTTPADVLYKGTITRIIGEDSPSLDRGSDSLPKGVITYEGKKHVLST 1281
 Db 461 A-----PPAAGVTSAPDTR---PAFG-----STAPPAHGV----- 487
 QY 1282 GGMSTVQCKEKDQSSSGPHEHAAKRTYDMMEGVSAGASISLEGMAIPERHS 1341
 Db 488 -----TSAPDTRPA---PGSTAP------ANGVTSAP-----PDRPA 516
 QY 1342 PHHLEKQHHRIGSIITQIGIPRSYVEAOEDYLREAKLKEEGTPPPP-----PSRDLTAY 1397
 Db 517 P-----GSTAP-PAHGVTSAPD-----TRPAGSTAPPAHGVTSAP 551
 QY 1398 KTOALGPLKUPAHEGLVATVKEAGRSIHETPREELRHTPELPLARPLKEGSIITQGTPL 1457
 Db 552 DTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPA- 597
 QY 1458 KYDTGASTGSKGADVRLIG---SPGRTPPVHPDVAADALERACYEISIKSRPST 1514
 Db 598 -----GSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTR- 639
 QY 1515 ASSSGSISARGAVIIVELCKPROSPLTYEDHGAAPAGHLPGSPVYTMREPTRLQESGI 1574
 Db 640 TAPPAHGV- SAPDTRPAAGST-----APPA-HGVTSAPDTR-PPA- 677
 QY 1575 SSSKASODKLTSTPREIAKSPHSYVPEHHPHISPEHLLRVSQVDLYRSIIPLAFP 1634
 Db 678 -GSTAPPAHGVTSAP-DTRPAAGSTAPPAHGVTSAPDTRPA- 727
 QY 1635 TSIPTGILDAAMAAYVLRHIAENPTYPHLYPYLLRGVYPTDALENQTIINDYITSSQ 1694
 Db 728 TSAPDTRP-----ABGSTAP-PAHGVTSAPDTRPAAGSTAPPAHGVTSAP 771
 QY 1695 MHNHTATAMAQRADMTLGLSPRESSLALNTAAGRGITIDLSQVPHLPVLPPTGTPATA 1754
 Db 772 ---DTRPAAGSTAPPAHGVTSAPDTR-----RPAAGSTAP 803

QY 1755 MDRLATIP-TAPQFSSRHSSPSLSPGPTH--LTKPTTSSSRERDRDRERERE 1811
 Db 804 AHGVTSAPDTRPA- -GSTAP- -PAHGVTSAPDTRPA- 837
 QY 1812 KSILSTTVHEHAPIMRPTGEGSGSGSGGSGSS--SRPASHAHQHSPISTPTOD 1869
 Db 838 ---GSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPAAGST- -PAHGVTS 889
 QY 1870 ALQORPSVLANHGMKGIITAVEPEKPTVLNSTSSSVRPAATPPATHCEPLGGLTLDGV 1929
 Db 890 APDTRPAAGSTAPPAHGVTSAPDTRP- -AGSTAPPAHGVTSAPDTRPAAGSTAPPAH 945
 QY 1930 PTLMEPVLPKEARVAVRPREPRADTGHAFKAPAR- -SGLEPA- -SSPSKSEP 1981
 Db 946 GVTSAAPDTRP- -AGSTAP- -PAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPA 997
 QY 1982 RPLVPVSGHATART- -PAKNLAPHN- -ASPD-PPAPASADPHREXTQSKPSTIQ 2034
 Db 998 GSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDNNRPAAGSTAPPAHVNVTASGASGS 1057
 QY 2035 ELEIRSLCYHSSYSPEGVEPV- -PVSSPSLTHDKGLPHHLELDKSHLEGELRPQP 2091
 Db 1058 ASTL- -VHNGTSARATTTTSPSKSTPSTPS- -HSDPTTL- -ASH- 1097
 QY 2092 GPVLCGSAHLPHLRPLPESQPSPLQTPAPGVKGHQRVTLAHISEVITODYTRHH 2151
 Db 1098 -STKIDASTHSTVPLTSSNHSSTPOLST- -GVS- -FFLSFHISNL- 1141
 QY 2152 PQQLSAPLPAPLYSPGASCPVLDLRPPSDLYL 2185
 Db 1142 --QFMSLEDPSTDY- -YQELORDISEMFL 1167

RESULT 77
 T45467
 collagen alpha 1(ii) chain precursor [imported] - horse
 N.Alternate names: type II collagen
 C.Species: Equus caballus (domestic horse)
 C.Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
 C.Accession: T45467
 C.Richardson, D.W.; Dodge, G.R.
 submitted to the EMBL data library, June 1996
 A.Description: Cloning of equine type II collagen and modulation of its expression in eq
 A.Reference number: 222977
 A.Accession: T45467
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: mRNA
 A.Residues: 1-1418 <RIC>
 A.Cross-references: EMBL:U62528; PDB:AAB05773.1
 C.Superfamily: collagen alpha 1(i) chain; fibrillar collagen carboxyl-terminal homology,

Query Match 2.5%; Score 335; DB 2; Length 1418;
 Best Local Similarity 22.4%; Pred. No. 0.00011;
 Matches 340; Conservative 93; Mismatches 567; Indels 520; Gaps 81;

QY 721 ABAALHAGNEV-----PRGECSPATVNSSDTESIPSPHTEAK-----DTGQNG-P 767
 Db 18 AAVLCKHGDVQAPQPKQKGRPDIXOIVGKPPRGQGRGEGGPRGDRDKGEKAP 77
 QY 768 KP-----PATLGADGP-PPGPTTP-----RTSRAP 794
 Db 78 GPRGRDGEFGTGNNGPFGPPGPGFIGNFAAQMAGFDEKAGAGMVGMPGM 137
 QY 795 EPTPASEATGATP-----PPAPSPSAPPVVPKEKEEETAAAPVEEGEQKPPAA 848
 Db 138 GRGPPGPAAGAPGQFQGNPGEPPGEGVSGMGR-----GPP-----GPPKPG- 183
 QY 849 EELAVDTGAEBEPVSECTEAEBSGPAKGDAAEAATAGALKAEKKEGSGRATTAKS 908
 Db 184 -----DDGEAGKPGMS-----GERGPPGQAGRFPPT- PGLPVYKGRGVGLDGAKGE 232
 QY 909 SGAPODSDSATCSADEVDABEGGDKRLSPRPSLLTPTGDDPRANASPOKPLDKOLK 968

Db 2041 -----TITATGTTNPESTPTGTTPIPVLTATTAATTAATSTVTPS--SALGTH----- 2088

Qy 2086 LRPQPGCVKLGGEAHLPHLRPLDEQOPSSPLQTR--FGVKGHQVVTLAQHISVYI 2143

Db 2089 -TPVP-----WTTATTHGRSLPPSSPHVTPTAMTSATGIIIG-----TTHITE-- 2131

Qy 2144 TQDTRHHHQGLSLAPLPLVSPGASCPVLDLRPRPDVLTLPDPDHGAPARGSPHSGG 2203

Db 2132 -----PSTGHTTTPAATGTTGTTOPSTPLSSPHPSRTTESPSPGTTTGG--HRTGT 2181

Qy 2204 KR-----SPENKTSVL-----GGGED 2220

Db 2182 SRTATATPSTKTRSTLLPSSPTAPITTVTTGCEQGCAMSEMDYSPMPGSGGDFD 2241

Qy 2221 -----GIF--PVSPEGKTEPCH--SRSAVYPLLYRQGO----- 2251

Db 2242 TYSNIRAGACVCEOPGLCECRAQKQPGVPLRELDQVVECLDPLGLVCRNEQVQKFKMC 2301

Qy 2252 -----TERSMGSKSGPNTSQPAFAFESKL-----ESNSAMVKSCKQ 2288

Db 2302 FNYEIRVFCNCYGHCPSPATSPATSPSTPGTTWILKLTATTTATSTGSTATPSSTQ 2361

Qy 2289 -----EINKKLNTNREBEYNIQPGTEIFNMPAITGTGLMTYSOAVOEHAS 2337

Db 2262 GPAGTPHVSTTATPTVTSKATP--FSSPG-----TATALPALSTATTPAT 2409

Qy 2338 TNNGLBAITRKALMGKYDQMBSPPLSANATPPLNANASLPAAMPITAADGRSDHT--L 2394

Db 2410 S--FTALPSSSLGTTWTR-----LSQTTTPMATMTSTATSPSTEIV-----HTSVL 2454

Qy 2395 TSPGGGAKXVSGRRSSKAKSPAGLASGDRPSPVS-----VHSEDDCKRRPPLNRY 2449

Db 2455 TT-----TATTTGATGATVATPSSSTPGLAHHTKVTPTTTGTTVPSSSPGTAATP--PV 2506

Qy 2450 W---EDRPSAGSTPFP-----YNPLMR-----LQAGVWASPPPGPLPAGSGP 2490

Db 2507 WISTTTTPTTSGTVPSPISPGTHTTPVLTTLTTPQVATGSMATPSSSTQTSQSTP 2562

RESULT 76

A35175

mucin 1 precursor, repetitive splice form A [validated] - human

N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; epistatlin

N:Content: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000

C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51

R:Litlenberg, M.J.L.; Vos, H.L.; Gemmelsen, A.M.C.; Hilkmann, J.

J. Biol. Chem. 265, 5573-5578, 1990

A:Title: Epistatlin, a carcinoma-associated mucin, is generated by a polymorphic gene enc

A:Reference number: A35175; MUID:9020294; PMID:2318825

A:Accession: A35175

A:Molecule type: mRNA

A:Residues: 1-952, 1033-1344 <LIG1>

A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA5804.1; PID:g182124; GB:

A:Experimental source: splice form A

A>Note: Genbank entries HUMEPIS1A1 and HUMEPIS1A2 present only the amino-and carboxyl-er

A:Accession: B35175

A:Molecule type: mRNA

A:Residues: 1-19, 29-952, 1033-1344 <LIG2>

A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA5806.1; PID:g182129; GB:

A:Experimental source: splice form B

A>Note: Genbank entries HUMEPIS1B1 and HUMEPIS1B2 present only the amino-and carboxyl-er

R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel

J. Biol. Chem. 265, 15286-15293, 1990

A:Title: Molecular cloning and expression of human tumor-associated polymorphic epithel

A:Reference number: A35886; MUID:90368715; PMID:1697589

A:Accession: A35886

A:Molecule type: mRNA

A:Status: not compared with conceptual translation

A:Residues: 1-19, 29-992, 1033-1344 <GEN>

A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870

A>Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence

R:Plan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.

J. Biol. Chem. 265, 15294-15299, 1990

A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.

A:Reference number: A35887; MUID:90368716; PMID:2394722

A:Accession: A35887

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19, 29-1109, 'S', 1111-1339, 'A', 1341-1344 <LAN>

A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599

A>Note: Genbank entry HUMPAMU containing four fewer copies of the tandemly repeated sequen

R:Wreschner, D.H.; Harevent, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.

Eur. J. Biochem. 189, 463-473, 1990

A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera

A:Reference number: S10571; MUID:90276413; PMID:2351132

A:Accession: S10572

A:Molecule type: mRNA

A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WRE>

A:Cross-references: EMBL:X52229; NID:937053

R:Wreschner, D.H.

submitted to the EMBL Data Library, March 1990

A:Reference number: S40293

A:Accession: S40293

A:Molecule type: mRNA

A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WR2

A:Cross-references: EMBL:X52229; NID:937053; PIDN:CA36478.1; PID:937054

R:Abde, M.; Siddiqui, J.; Kule, D.

Biochem. Biophys. Res. Commun. 165, 644-649, 1989

A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated

A:Reference number: A36735; MUID:90088473; PMID:2597151

A:Accession: A36735

A:Molecule type: mRNA

A:Residues: 1-142, 'Q', 144-162, 'Q', 164-168 <ABE>

A:Cross-references: EMBL:M31823; NID:9181542; PIDN:AAA35757.1; PID:9181543

R:Maizawa, Y.; Miyauchi, T.; Hamanouchi, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H

J. Biochem. 112, 609-615, 1992

A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu

A:Reference number: JX0235; MUID:91212189; PMID:1478919

A:Accession: PX0066

A:Molecule type: mRNA

A:Residues: 998-1011, 'ES', 1014-1017, 1018-1032, 'T', 1034-1037, 1038-1057 <MAS>

A:Experimental source: gastric carcinoma cell

R:Zirhan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.

FEBS Lett. 356, 130-136, 1994

A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r

A:Reference number: S51026; MUID:95080414; PMID:7988707

A:Contents: annotation

A:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c

partial repeats. The repeat shown is defined by SmaI nuclease sites.

C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly

C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S46146

C:Genetics:

A:Gene: GDB:MUC1; PUM

A:Cross-references: GDB:120705; OMIM:158340

A:Map position: 1621-1623

A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3

C:Superfamily: polymorphic epithelial mucin

K:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphi

F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>

F:1-62/Region: mucin 1 amino-terminal non-repetitive

F:1-33/Domain: signal sequence #link PREA #status predicted <SIGA>

F:1-19, 29-33/Domain: signal sequence #link PREB #status predicted <SIGB>

F:1-19, 29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>

F:1-19, 29-212, 1033-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>

F:133-1017/Region: 20-residue repeats (GSTAPAHGVTSAPDRPAP)

F:133-1344/Region: mucin 1 carboxyl-terminal non-repetitive

F:1245-1272/Domain: transmembrane #status predicted <TM>

F:1046, 1064, 1118, 1144, 1222/Binding site: carboxylate (Asn) (covalent) #status predicted

F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 2.5%; Score 335; DB 1; Length 1344;

Best Local Similarity 21.5%; Pred. No. 0.0001;

1412 -----EGVATVKEAGRSIHIEIPREELHTELPRLAPRLKESITOGTPIKY 1459
DB PGASAGASGGEGVSARPSOGG-----TPSGTPASAPVAAGPAGGE----- 599
1460 DTGASTGSKKHDVRLIGS-----PGRTPPPHPLDV-----MADARALE 1500
DB GSGSGTSGS-----GSGAPMAARPGGGSDAPVDLGLTSGRGGSGAGNGAGTIGE 650
1501 RACYESLKRPPGTASSGSGIARFAPVIVPELGRKROSPLTYEDHG-----APPAGHLP 1555
DB 651 TGAAGTGS-ASRPGSGGTGNGSGTGVPASGKGTGSGTGNGGAGGAGVAV 709
1556 RGSPTVMBEPTPRLGSLSSSKASQDKLITP-----REIAKSPHSTVPEHHPI 1608
DB 710 RGG-----GGAGGAGNETRPLTCTVVVDVRLGLKLESAMTSFVMDSAGNOV 756
1609 SPEYHLRGVS 1619
DB 757 WPDALIKGVA 767

RESULT 75

T45025
mucin MUC5B, tracheobronchial (imported) - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #ext_change 21-Jul-2000
C/Accession: T45025
R/Accession: J.L.; Guyonnet-Duprat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A>Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A/Reference number: Z2899; MIM:19716151; PMID:9013550
A/Accession: T45025
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3570 <DES>
A/Cross-references: EMBL:Z72496; NID:G1934502; PIND:CA9657.1; PID:G1834503
A/Experimental source: placenta
C/Genetics:
A/Gene: MUC5B

Query Match 2.5%; Score 335.5; DB 2; Length 3570;
Best local Similarity 19.5%; Pred. No. 0.00028;
Matches 409; Conservative 203; Mismatches 835; Indels 649; Gaps 91;

687 PAASEBAAPPVVEDEMAAGVSGNBEEMVEBAALHNSGNEVREGSGAPATVNNSS 746
DB PPSSTPTTPIPPVLTATTPAATST-----VTPSSAL-----GTHHTPVPRVTRA 870
747 DTR-----SIPSPH-----TEAADTONGPKPATLGA-----DGRPP 780
DB 871 TTHGRSLSPSPHVTCTAMTSATGILGTHHTERSTGSHT-----PATTTGTOHSTPAL 927
781 GPTTPRTSRAPTEP-----TPASEATGATTP-----PAPPSAPAPPVPRK 824
DB 928 SSPHSRTTSPSPSGTTPGHTTATSRITATPRTKRTSTLLPQPT-SAPITTVT 986
825 EKEKEETAAAPVVEEGEQKRAAEEL-----AVDTGAEBPVKSECTEAEKEGPAK 877
DB 987 MGCPQCMSEWLDYSTPMGPSSGDDPTTYSNIAAGAVCEQJLGECAQAOQPGVPLR 1046
878 KDAEAETAGALKAKEKEG-----GSGRATTAKSGAPQSDSS 918
DB 1047 ELQGVVECSLDLGLVCRNREQVKFQMCFNVEIRVFCNNGHCSPTATSTATPSSTPG 1106
919 ATCSADEVDEAEGDKNLLSPRSLTLTPGDPRAANSPOKPELDLKLQKRAAALPIQV 978
DB 1107 TTM-----ILTELTATTTSTSTSTATPSTL-----RTAPPPVLT 1144
979 TKVHEPPREDAPTKAPAPPPONLQESDAPQOQSSSPGRKSRPAPPADKEAPAAE 1038
DB 1145 TTATTP-----TTSKATPSSSBGTATLP-----ALRSTATTPTATSVPIPISSSLGT----- 1194
1039 AOKLPGDPPCWTSGLPFVPVPREVIKASPHADPSAF-----SYAPPGH 1082

1195 -----TWTLSSQTTPTATMSTATPSSTPETAHTSTVLATATTTGATGSVATPS- 1244
DB PLPLGHDTPARVPL-----PRPPTSNPPLISSAKHNSVLEROIAGLSQMSV- 1131
1083 STPGTAHTTKVPTTTTGTPTATPSSSPGATLPVWISTTTPT-----TRGSIVT 1295
1245 QLVHVPSEHAKAPVPATVMTLPLPMDPKLAPSGVQEOQLSPGOAGPESLGVPTAOE 1191
DB 1132 PPSITPGTHATVLTITTTTVA-----TGSNATPSSSTQISGTPSL---TTTA 1341
1296 ASVLKGTALGSPGSIITKGI-PSRVPSDAITYRGSITHTGPADVLKGTITRIIGED 1250
DB 1192 TTI-----TATSGTTPSGTGRPIPVLTATTPAATSTVTPSSAL---GT-THRPVP 1395
1342 SPSRLDGRBESLPKHVITIEGKGNHLSYEGMSVTOCKEGRSSGPHETAAPKRT 1310
DB 1251 NTATVTHGRSLSPSPHVT-----RTAMTSATSGTLGTHITE---PSTGSHPAATGT 1448
1396 YDMMEGRVGRASIASIEGLMGRALP-PERHSPHNLKHOHNRISITQGI PRSVVEAOED 1369
DB 1311 TQ-----HSTPLSSPHSPSRSTESPPSPGTTTGGH-----TTATSRTATATPS 1493
1449 YLRREAKILKRETPPPPPSRDITEA-----YKQALGPILKPKPHNEG 1413
DB 1370 KRTSTILL-----PSSPTSAPIITVTVWGCSEPOCAMSEWLDYSMPGP---SGDFFD 1543
1494 LVATVKEAGRSIHIEIPREELHTELPRLAPR-----PLKEGSIITOGTPLKYDTGAST 1466
DB 1414 TYSNIRAAAGVAVCOF-----LGLCECAQAOQPGVPLRE-LQGVVECSLDLGLVCR 1592
1544 GSKKHVDVSLIGSDPRTPPVPHPLDVMAARALRACYESLK-----SRGTAS 1516
DB 1467 N-----REQYCK-----FKMCFNVEIRVFCNNGHCSPTATPS 1626
1593 SSGSANGAPVIVPELGRKROSPLTYEDHGAPRA---GHLPGRSPVMBEPTPRLGSL 1574
DB 1517 TATPSTPGTWTWLEOTTAATTAATGTAIPSTPETAAPPKVLSQATTP-----TA 1681
1627 SSSKA-----SOPKLTSTPREIAKSPHSTVPEHHPIPSYEHLLRGVGVLDYRSHPLA 1631
DB 1575 TSSKATSSSSRPTATTLVLNSTATKSTATSTPISSTL-----GTGTGQNBPRHMA 1736
1682 FDPSTIPGCIPLDAAAYVLPRLHAPNPTVPHLPVILRGVPPDPALENNQTIINDVIT 1691
DB 1632 TWTST-----HPSST-----PETT----- 1750
1737 SQQMHNTATMAORADMLRGLSPRESSIALNVAGPRGIIDLSQVPHLPVLPPTGTP 1751
DB 1692 HSTVLTATTAATTAATPSSTPS-----TPGTT 1779
1751 1752 -----ATMDRLAYLPTAPOFSSRHSSSPISPGCPHLLKPTTSSSEBRDRDRD 1805
DB 1780 WITELTATTAATTAALP-----HGPPSTPGTWTWLEPSTTAATVTVFGSTATAS 1830
1806 RDREREKS---ILSTTVYHARITWRPTQSGSSGSGSGGSGSSSPASHAHQHSPI 1863
DB 1831 STRATAGLKVLTATTP-----TVISERATPS-SEGTATALALAKSTA----- 1875
1864 SPRTDALQORPSVLNMGKGIITAVEPSKPTVLKSTSTSSPVPAATFPATHCPLQG 1923
DB 1876 TPTATNSVTAIPSSSLGTAMTLSQTTT-----TATVSTATPSSPT-PEVATSTVL 1926
1924 TLQGVPTIMEPVLLPKAPRVARPERPRADTGHAFKAPRASC-----LBPASSP 1975
DB 1927 TTAT-TTRTGSVATPSSSTGTATHTTKVPTTTTGTATPSSSPGATLPVWISTTTT 1985
1976 -SKGSEPRVLPVNSH-ATIAKTPAKNLARNHNSDPPAPASADPRREKTOGKPEFI 2033
DB 1986 TTRGSTVTPSSIPGTHATVLTITTTTVAATGSMATPSSSTQISGTPSLTTAT----- 2040
2034 QETELRSIGVHGSSVPE--GVEPVSPV-----SSPLTHDKGLPKHLELDKSHLGE 2085


```

Db 310 DFTA-----KMIEELKAKS-----RAPLVETKPVKMPGEGIHKEFKTESQFGLNST 355
Qy 609 LNEBSRWTEEMET-----AKKGLLHGKRWMSIAAMWGSKTVSOCKNFTYKRN 661
Db 356 KIHNRQPTSDFTETIAQTAEKRL--GAQAGASEPSSS-----SIHKKKKT 406
Qy 662 LDEILQOKLME--KERNARKK-----KABAASEEAAPPVEDEMEASG 709
Db 407 VRE-LQTEMKGESYKKAARAKASAGRSQAPAPAPASELQDPDQGLMSDPG 465
Qy 710 VSGNEEVEEABALHASGNEVPRGCSGPATVNNSDTESIPSHTEAANDTGNGPK 769
Db 466 SDSETEHQOKHPIPAVYTRRSARLSALPVTPKASSSSKMPPPSPSPGGRGRRP 525
Qy 770 -----PATLGADGPPPPPTPPRTSRAPITPT--PASEATGATP--PPAPSPS 816
Db 526 RTLSTMSNEPA---AAAVTPAPRGPRRSRAKVSENTEPLSEASAPVKRGRGRSRS 582
Qy 817 APPVVPKKEKEETFA--APPVEGEOKPPAAEELAVDTGKAEPPVK--SECTEAEBSG 873
Db 583 TMSITDESEPTSTTAARKSKRASEDEBE---ODLKITNKSPEKPKPKTTEETVGD 638
Qy 874 PAKGDAEAEATNE-----GALKAEKKGSGRATTAKSSGAPQDSDSATCSADEVD 927
Db 639 VLKRLRDTAKTTATVHTPGPLTRKME--RMRAPAVYS-SKKEKPKVAGSADSSIN 695
Qy 928 EAEQDKRRL-----SPR-----PSLTPGDP 951
Db 696 EEEHEDETMIEEQTLDPQOTSQOEPRISCGSELLDEQFDSSEHSGTVSAPRLTNP 755
Qy 952 -----RAVNASPOK--PLDLKQKORAAIPIQVTKVHEPREDAAPKAPAPAP 1000
Db 756 APPVEASEASAPKIDIPBOATPILALALALPVSPTAL--EPKAOENPFAELPTSE 814
Qy 1001 ----PPQNIQESDAPQOPGS-----SPRGSRSAPAPADKAEFAAEQK 1041
Db 815 ISGRAPQALPTSSQPTSGSAAAPVDLLEILSGAKTTTRKAPAPAVGSSISTTQQ 874
Qy 1042 LRGDPCTSGLPFVPPREVIVKASPHAPDPSAFYAPP-----GHLPLGLHDT 1091
Db 875 AP-----PTSVQAP--PTSCSAAPVDDLSEILSGAKTTTKTT 913
Qy 1092 ARPVLPREPTISNPPLISSA-----KHP--SVLEROIQA--IS-----QQM 1129
Db 914 KMPFVDQKKISSAPISDSAPTSHOQTPKSPKQILNSKYLDISDEDEEEBEEKOM 973
Qy 1130 SVQLHVPYSEHAKAFVGVTMGL-----PLPMDPKLAPSGVQKQOLSPRQAGPPE 1182
Db 974 EI-----VEEBEAP--PLSDSLQASBPSTATVPEKVAVVKLFSPEDISTVEAPPE 1026
Qy 1183 SLGVTQAQASVLRGTAL--GSVPGGISTTKGIPSTR--VPSDSATYTRGSIHTGPADV 1238
Db 1027 ASVPKAPATKIDQLADQASBEPKARKLPIARKIPKIKISLAPSSSTSDDL 1086
Qy 1239 YKGTITRIIGEDSPRLDGRGREDLPRKHVIVYEGKKHVLSEYEGASVTQCKEDGRSS 1298
Db 1087 MSDII-----AAKTKTKTK-----PK 1103
Qy 1299 GPPHETAPKRTYDMMEGRVGRALISSASIEGLMGRAPPERHSPHLKEOHIRGSI 1358
Db 1104 APPTVAQTRTKNLAKR-----KASPTPAGTTAPKQ--YIKKSIDSV 1147
Qy 1359 IPRSVYEAQEDYLREAKILKREGTPPPPPSRDLTEAYK----- 1398
Db 1148 LPPSSSSSTE-----PPSAPDSASTTSSMKKGGAIMEAVPCRGKAI 1192
Qy 1399 TQALPLKLPKRAHEGLVATVKEAGSIHIREBELRHTPELPL-----APPLKESINTG 1454
Db 1193 KREQPIGK--EMMVQNEKGGKKVVKI-KTHLRQALDLKIPPEBELKRPHEEKIKLG 1248

```

RESULT 74
F75518

hypothetical protein - *Deinococcus radiodurans* (strain R1)
 C:Species: *Deinococcus radiodurans*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: F75518
 R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.U.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma-
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
 A:Reference number: A75250; MIMD:20036896; PMID:10567266
 A:Accession: F75518
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-839 <WHI>
 A:Cross-references: GB:AE001904; GB:AE000513; NID:96458129; PIDN:AA10038.1; PID:9645814
 A:Experimental source: strain R1
 A:Gene: DR0458
 A:Map position: 1
 C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology

Query Match 2.5%; Score 336; DB 2; Length 839;
 Best Local Similarity 21.2%; Pred. No. 5.6e-05;
 Matches 206; Conservative 71; Mismatches 342; Indels 352; Gaps 34;

```

Qy 730 EPRGCEGSPATVNNSSDTEISPSHTEAANDTGNGPKPATLGADGPPGPTPPRT 789
Db 68 EVPSNNAAPLSTPSPLASPKKPAATKPTPT---PAKATPAPEPKPEPTPEPK 122
Qy 790 SRAPLEPTPASEATGAPTPPA-----PPSPAPPPVVPKKEKEETAAAPVEBEGOKP 845
Db 123 PETPEPPKPAPEPKPPTPAPELKPVPVQOTPEPVTPKVPTEVTPKPAPTPEVQL 182
Qy 846 PAEEELAVDTGKAEPEVKSCTEAEBSGPAKQDAEAAEATBEGALKAEKKGSGRATT 905
Db 183 FVAQTPPV---AKPPV----- 195
Qy 906 AKSSGAPQDSDSATCSADEVDEAGCKNKLSPRSLTPGTGPPRANASPOKRLDKQ 965
Db 196 -----PAPTSQTP----- 204
Qy 966 LKORAAIPIQVTKVHEPREDAAPTPKAPAPPPONLQESDAP--QQPGSS----- 1018
Db 205 ----PPVQAPATRTTPPQAARPTNAAQTPAQAQTPAQAQTPAQAQTPATP 256
Qy 1019 ----PRGKRSAP--PADYEAFAA---EAQKLPDPPCTSGLPFPPVPPREV--IKA 1065
Db 257 AAPAQRPAAGCAPSPAPAPQAQANAPAGSVPAATVPESTPPAPSAQTPPTRETAAQTEA 316
Qy 1066 SPHADPSAFYAPRGHPLPLGLHDTARPVLPRTTISNPPLISSAHPSVLERQIGAI 1125
Db 317 SPAAPNSA---AAPEP-----ASEPVAGRPTAASPEBSAS----- 351
Qy 1126 SQGMSVQLHVPYSEHAKAPVGVTMGLPLPMDPKLAPFGVQKQOLSPRQ-----A 1178
Db 352 -----PVT-----VTPRGETPDTAASA 368
Qy 1179 GPPESLGVPTQAQASVLRGTALGSVPG--GSITTKGIPSTRVPSDSATYTRGSIHTGPADV 1237
Db 369 GTPSAGRVTPAPAPASGASAPAPPGAGSQTPPIPATPIPATPAGRSSG--SAGTAA-- 425
Qy 1238 LYKGTITRIIGEDGSRLLDRGREDLPRKHVIVYEGKKHVLSEYEGMSVTQCKEDGRSS 1297
Db 426 ----ARPNAAPAVSSEDSVSGLPF-----REDAPAE 454
Qy 1298 GPPHETAPKRTYDMMEGRVGRAL--SSASIEGLMGRAPPERHSPHLKEOHIRGSI 1355
Db 455 SSP--VAASPARGASSASAPAAAVPRAVSG--GSVSAPTPTAPVAEQGVPPSP 510
Qy 1356 TQIPRSYVEAQEDYLREAKILKREGTPPPPPSRDLTEAYKTOALG--PLKAKPH-- 1411
Db 511 SAAAPRGASS-----AAPASAPAAARGSGAAGAGASAPAAAPAPQ 556

```

```

QY 817 -APPVPEKEKEBTAAAPVEEGE-----QKPPAAEELAVDTGKAEVYKSECT 867
Db 221 GARPGE-----GQFQGNPEGESGAGCGMGRPGPGSGKPDGDEAGKPKGS--- 268
QY 868 EBAERBPACKCAEAALAT-----ABGALKAEKKGSGGRATTASSGAPQDSATCS 922
Db 269 --GEGRPPOGARGPGPTGRLPGVKGHRGYPGLDGKGEAGAA--GAKGSGASGAG 323
QY 923 ADEVEBAEGDKNRLSPRSLTPTGDRANASPOKPLDKLKORAAATPPIQVTKVH 982
Db 324 A-----PCGMGRGLPBERG----- 338
QY 983 EPPREDAPTK-----PAPRAPPONTLPESDAPQPGS-----SPRG 1021
Db 339 RCGSSGAGARGNDGLPGAPPGPVGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 398
QY 1022 KSRSPAPRPAKFAAEOKLPDPRCWTSGLPFPVPRREVIKASPHADPSAFSVA-- 1079
Db 399 ESGTPGSGPSPSGA-----SGNP--GTDGIP-----GAKGSGASGAGIAGAPG 438
QY 1080 PGHPLPLGLHDTAPVLPRLPTISNPPLISSAKHPSVLEROIGALISQMSVOLHVPSY 1139
Db 439 PGPGRPGPGAGAPGLGPKGCT--GDPGLAGPKG-----HGPKGE 477
QY 1140 HAKAPVGVYTMGLPLM-----DRKCLAPSGVAKOBOLSPRGAGPESLGVPTA 1189
Db 478 IGSA--GP--OGAPGAPGEGKKGARGEPGAAP-----LGPBERGAGAPGAPG 525
QY 1190 QEASVLRGT-----ALGSPGSGSLTKGIP--STRVPSDAITVRSGITHTPADVLYK 1241
Db 526 DGLAGKRGARGENGVLGPKKGNDDPRGPRGLRGARGLT-----GRPADAGPG 578
QY 1242 TI--TRIGED-----SPSRLDGRGEDSL--PYGHVLYBEGKCHVLYSGMSVTOC 1289
Db 579 KVGPSGAAGDGRPPRGPGA--RQPGVWGFPPRGANGEPKAGE--KGLGAPGLRGL 636
QY 1290 SKEDGSSSGPRHETAP-----KRTYDMEGNVR-----AIGSASI 1327
Db 637 PGDGETGAGGPNRGPAGEGEBORPPSGFGLRGPPGSGEGSGKGDQVPEBA 696
QY 1328 EGLMG-----RAIPRERHSPNHLKEONHIRSGITOGIPRSVVEAQEDYLREAKLIREGT 1383
Db 697 AGLVGRGERGFRGERS-----SGRGLQGS--RGLP-----GT 729
QY 1384 PPPPPSRDLTEAVYKTALGRLKRAHGLVATVKEAGRSYHEIRREBELRHTPELPLAP 1443
Db 730 PGTDGPKGATGSPGNGAOGP-----PGLQMPBERGAAGIS--GPGKMGDTGEGKPEG 782
QY 1444 RPLKEGSLTGTPLKVTGASTTSGKNDVRSLLGSGRTFPPVHPLDVADARALERAC 1503
Db 783 APGKDGSS-----RGLTGP-----LGRPG--PAGP----- 804
QY 1504 YEESLKRPGTASSSGSIARGAVIYBELKRPSPFLYEDHAPAGHLPRGSPVTKR 1563
Db 805 --NGEKESGPPSGPGLVARGARGDGENGPP--GP-----AGPAG--PPGS----- 846
QY 1564 EPTPLRLOEGSLSSSKASQDKLSTPREIAKSHISTYREHNPRLISPYHLLANGSVGLD 1623
Db 847 -----DQO--AGLKPDQGE--SGQKDGAGAPRPGSAPRQGP-----TGVMG-- 887
QY 1624 YRSHIPLADPTSLPRGILPLDAAAYLPRHLAPNPTYPRLVPRYLIRGYRDTALLENRO 1683
Db 888 -----PKAGAGAGAPPAATGPRGA--GAVGTGPR--GNRGP-----GPPGSGKKGPK 934
QY 1684 TIINDVYTSQOMH--HNTATAMAGRAM--LKGISPPRESSIALNYAA 1726
Db 935 GVRDAPGPTGACPGIAGGAPAGGEGEPEDGSPGDPSPGQISGKRGIVGLRGOR 994
QY 1727 GPRGILDSQVPH--PVLVPRPTGTALMDRLAFLYTAAPRPSSSHSSPLSPGSPTH 1784
Db 995 GERGP-----PGLRPSGEPKKGKGGSSGDR-----GPRPVPVPG 1031
QY 1785 LTKPTTSSSERERDRDRDRERDRERKSLITSTTV--EHAPIWRPTEOSSGSSG 1839

```

```

Db 1032 LTGPAGETGREGNDGSGPPGRD-----GATGKEDRGHTGPIGAGAPGSPGAPG 1082
QY 1840 SSGGGGSSSRPASHA-----HOHSDISPTODA-----LQRPSTLHN--TG 1882
Db 1083 PVGPPTGKQDGRGEGSGGPIGPGGAPAGRLNGPQGRGKGEANGENGEGOKHRRFTG 1142
QY 1883 MKGIITAVEBSPKIVL--ASTSTSPVRPAATTPRATHCPLOGTLGQVYPTIMEPVLLP 1939
Db 1143 LQGL-----PGPPTAGDQAGTSGAGSPRG--PPGPVPSG--KDG-----S 1182
QY 1940 KEAPRVARPERPRADTGHAFIAPKAPASGLEPASPSPSGSEPRELVPSGHATIAPTA 1999
Db 1183 NGITGPIGPSPRPRRSGEITGAPGPGSG-----PPGPRGPRG--IDWSAF 1228
QY 2000 KNLAPHHASDP-----PAPPASAS 2019
Db 1229 AGLSQPEKGPDPMRMYRADQASSS 1252

```

RESULT 73

```

126517
Hypothetical protein Y18D10A.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26517
R:Haritz, B.
Submitted to the EMBL Data Library, December 1998
A:Reference number: Z20226
A:Accession: T26517
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1634 <WIL>
A:Cross-references: EMBL:AL034393; PDB:CAA22308.1; CESP:Y18D10A.1
A:Experimental source: clone Y18D10A
C:Genetics:
A:Gene: CESP:Y18D10A.1
A:introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2; 1

```

Query Match 2.6%; Score 337; DB 2; Length 1634;

Best Local Similarity 19.7%; Pred. No. 0.00011;

Matches 295; Conservative 181; Mismatches 530; Indels 494; Gaps 64;

```

QY 172 LTONMDRV-----DREITVWEQISKLK-----KKQQLLEBAKPEPEKVPSPPT 219
Db 26 LMQNNQRAEMQDPFKVQAECEVLAMKLVTDAVRESVQREBAEAINAK----- 74
QY 220 ESKHRSIVQIYDENRKAABAAHRLLEGIGQVVELPYNPSPDROYENIKINQAMRK 279
Db 75 ETAEASLKQGYDSSEFYKKSA-----SLQEQDDAEKHKQO----- 111
QY 280 LILVFRNRHAKQKQKFCQRYDQMEALKEKVERIENNPRAKESKVRVEYKOPPE 339
Db 112 -----ANYRANK-----YDSVC-AL--TRLEANN-----ID 135
QY 340 IRKRELQERMQSRVQSGSGL-----SMSAARSH-----EVSEIIDGLSEQ 382
Db 136 IKSKESTYKLEIETGKEMTSLDNQVTTLLKAKINSSKTKLDYLYKSVIQVDALST 195
QY 383 ENLEKQMRGLAVIPMLYADQORIKFIMNGIMADPMKYVQROVNMWMSHQEETPRE 442
Db 196 SNIKQSKSI-----KDLVIVR-----RDVRL 219
QY 443 KFWQHPKNGLIASFLEKTVACVLYYLTKKNENYSLVRSYRRGKSQOQOQOQO 502
Db 220 FNRSGP-----WKSLSTSPSISATBAEASEDPE 249
QY 503 QOQOQOQOQMP-----RSSQEKQEKQKAEKEENP--EVENDKEDILKEXT----- 550
Db 250 HLDVITTEPAPRIGEQTLSDDEEBEVEPEDEADDEVREINQLENELLQRYKEKTVRRP 309
QY 551 DDTSGGDNDEKAVASGKRTANSQGRRKRI--TRSMNEANSSEALTPQOASIASME 608

```

```

Db 734 GIPLEPTDA-----SGA-----VIGPDGELIIPDASGKPLSADGSPLEPTDNNGNVYLVPADE 7984
Qy 1208 ITKGIPTSRVPSDAIYTRGSIITGTCTPADVLKGTITPIIGEDSPSLDRGRSDSLPGH 1267
Db 7985 VTTKV-----LPTDSD-----GNVYH-----PIRPPD-----TFLGTADGSPFTDDGQ 8024
Qy 1268 VIVECKKHVLSYEGMGVNTQCSKEDGRSSGPPHE-----TAAPKTYMMGKRGVRAIS- 1323
Db 8025 AIEKODEKPIGPDGQVLPTDASGNVYIPVIGPDQALPTDCKGKTYVPVPGDPTPLST 8084
Qy 1324 --SASIEGIMGRAIPPEHNSPHHLKEQHIRGSIITQGIIPRVEAEDYLREALTLRE 1381
Db 8085 DASGLILGDEPIPTD-----ASGKPLSAD 8110
Qy 1382 GTPPPPPSRDLTEAKYTOALGPLLKAHGLVAT--VKEGSIHEIPREELRHTEPL 1439
Db 8111 GSPLP-----TDAVGNVYILVPSDDGVIRTHPTDESGNTIYITKPD----- 8151
Qy 1440 PLAPRLKEGSIITGTPKPD--TGASTGS-----KKHVRSLIGSPGRTPP-----VH 1488
Db 8152 -----GTPLATSTGAFTVDDGOVITEKDEKPIGPDQVLPTDASGNVY 8197
Qy 1489 PLDVADARAL-----ERACEESLKSRRPGTAS----- 1516
Db 8198 PV-TSSDQVLPTDAKPIVVDGSKPLPTDASGNVYDNNKPIYIEGEEPTGPDQKLS 8256
Qy 1517 -----SSGGSITAR-----GAPVIVPEL 1533
Db 8257 KNEKEGWYPLVDFKGFVEITDNDKPVITVVDNNGNELSKNDGNMIDLSNEIDTDEL 8316
Qy 1534 GKPROS-----PLTYEDHG-----APAGHLPRGSPVMREBTPRLQGS 1573
Db 8317 GRPLDSGNPKKFDGNHNVIAPOIEEBETTPAIFL--IIDGPIEDDDVYTDKGN 8374
Qy 1574 LSSSKA-----SQDRKLTSTPREIAKSPHSTV--PEHHPHPIPSYEHLLRG 1617
Db 8375 VIPTNSBEKPIDENGQVLPKNEDGFVK-PKE-ADTTGSTIYSPGSLPTD----- 8424
Qy 1618 VSGVDLYNSHIFLAFDPTSIP-----KGIPLDAAAYY-----LPHILANPTY 1661
Db 8425 ASGAATIGDGEPIPTDSSGKPLSKDGSPLPTDASGNVYILVPSGEGVTDLSPTDEAGNTIY 8484
Qy 1662 PHLYPVLIRGPDIAL--ENKQIILINDYITSGQMHNNTAAMQARADMKGSPRSS 1719
Db 8485 PITKPDGTLATDSTGTFVADGQIIEKX-----DEGKPI----- 8519
Qy 1720 LALNYAAGPRGIT--DLQVPHLPLVLPPTGPTATMDRLAYLPTAPQ--PFSSRHSS 1774
Db 8520 -----GPDQGVLPDASGNVYIVIGPDQALPTDSSGKTYVPVVRGPDGTPPLPTDASG 8572
Qy 1775 SPLSGG--PTHLT-KPTTSSSERERDRER--DRDREREKSIILTSTT-TVEHAPI 1826
Db 8573 AVITGDGPIPTDPSGKPLSADGSKLPTDINNVYLVPADEVYTKVLPTDSDGNVYH-PI 8631
Qy 1827 WRP-CTEGSSSGSSGSSGSSSRPASHSHAHQSPISPRQDMLQGRPSVLANHTWKMG 1885
Db 8632 TRPDSTPLGTDSG-----SFTTEGQIVKNDG--Kp-----IGPDG 8668
Qy 1886 ITTAVEPSKPTVLASTSTSSPVRPATPPTATCPGLGTLGVYPLTMEPVLLPKEAPRV 1945
Db 8669 QVLPTDSSDNYIYPIGSDQAMPDT-----TGSVYIYLVPSDGVIGSPKV 8717
Qy 1946 ARPEPRADTGHAFLAKPPASGLERPASSPKGSEPRPLVPVSGHATITARPANLAPH 2005
Db 8718 AKP-----VGRPGKVLPTDASG-----H 8735
Qy 2006 HASPPRAPASAPDPRHEKTSKPSIQELRLSLGYHGSYS-----PEGVEVSPVS 2060
Db 8736 FIGPGPIF--TD-----YGVTVSDTVTTTDDI-----PLS 8764
Qy 2061 SPS-----LTHDKGLPKHLEELDKSHLEGLRPK-----QPGFVKLGGEAA 2101

```

```

Db 8765 NDSTGAFITEDGTIENNEDGKPIGPDQVLPTDAGNVYIPALPGDQALPTDESGNV 8824
Qy 2102 HL-----PHLRPLRES-----QPSSSLQOTAPGV-----KGQVYVTLAQH 2138
Db 8825 YVVRGPDGTPPLTDSGAVIGPDDEPIPTDASGKPLSADGSPLEPTDNNGNVYLVPADEV 8884
Qy 2139 ISEVITQDYTHHHPOQLSAPLAPL-----YSF-----PGASCPVLDLRP--PSDLYLPP 2187
Db 8885 TTKVLPTDSSGNVYHPIITRPDGTPLGTDSGSPFRDDQALAEKDEKPIGPDQVLPTD 8944
Qy 2188 -----PDHGAPARGSPHSEGGKRSPEPKNTSVLGGEGDIEFVSPPEGKTEFGHSRSA 2240
Db 8945 ASGNVYIPVIGPDQALPTDESGKT-----YVVRGPDGTPPLTDSGA 8988
Qy 2241 VYPLLYRDGEQTESRNGSKSPGNTSQPRAFSKLTESNSAMVSKKQEIINK-LNTHNR 2299
Db 8989 VIGL--DGEPIPTDASGKPLSAGSPLE-----TNNNGVYLVPADEVYTKVLPTDSD 9039
Qy 2300 NEPEYNISQPGTEIFNMPALITGTLMTYRS-----QAVQEHASTMGLAIIIRKALM 2351
Db 9040 GNVYHPIITP-----DGTPLGTDSGSPFRDDQALAEK----- 9073
Qy 2352 GKTDQWESSPPLSANAFNPINASASLPAAMPITADGRSDHTLTSPGCGGAKV-----S 2406
Db 9074 -----DEGKPIGPDG--QVLPTDASGNVYIPVIGPDQALPTDSDS-----GKTYVPVRGPD 9122
Qy 2407 GRPSSRKAKSPAPPLAGDGRPPSVSVHSEBDCNRRTPLTRWEDRPSAGSTPFVYN- 2465
Db 9123 GTPLEPTDASGAVIL--DGEPIPTDAS-----GKPLSAGSPLEPTDN 9162
Qy 2466 -----PLMRLOAGVNASP--PPGLPAG--SGPLA--GPHHAWDEBPK 2503
Db 9163 NGNVYVPAHNVHTTKVLPTDSSGNVYHPIITRPDGTPLGTDSGSPFRDDQALAEKDDGK 9222
Qy 2504 PL 2505
Db 9223 PI 9224

RESULT 72
A0333
collagen alpha 1(I) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C:Accession: A40333
R:Su. M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A:Reference number: A40333; MUID:92011898; PMID:1918153
A:Accession: A40333
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1492 <SUA>
A:Cross-references: GB:M63596
A>Note: this sequence is presented as substitutions relative to another sequence in a fi
e they replace: the appropriate interpretation of the sequence figure was reconstructed
C:Superfamily: collagen alpha 1(I) chain, fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 2.6%; Score 337; DB 2; Length 1492;
Best Local Similarity 22.3%; Pred. No. 9.6e-05;
Matches 327; Conservative 99; Mismatches 556; Indels 482; Gaps 74;

Qy 726 ASGNEVPRGEGSGPATYNNSSDTEISIPSPHTEAK-----DTGQNGPK-----PP 770
Db 101 SSGGQVAKGGEPRDIDVVGPRGPQGPQSPGEGPRGGRGKGTGAGPGRDGEPP 160
Qy 771 ATLGAQGP--PPGPTTPRTTSRAPIEPTPA--SEATCA-----PPTPPAPPSPS 816
Db 161 GTPGNPGAGPGPGPGPGLGAGNFAQMTGTFDEKAGAGQMGVMOGPMGPGPGPGPPS 220

```

Db	1394	POPRSRUKA-----DTEEFLEFRKOTPSAGAMH-----TP	1426
Oy	1444	RPL-----KESGIPOGTP-LKYDTGASTYGSKH-----DVRSLIGSPGR	1483
Db	1427	KPAVGEKDOIINTPLGTVPQKLDROGNLPGSNRRLQTRKEKAQALBELTGRLEFORPCTD	1486
Oy	1484	FPV-----HPLDVMA-----PALERACYESLSKRPCTASSGGGSLA	1523
Db	1487	NPTADEKTTKILCKSPQSDPADPTNTKORPKKSLKKAUVEEFLEFRKLTPSAGAMH	1546
Oy	1524	RGAPVLYPE-----LCKPRQSPLYTEDHCAFRFGHLPQSPVMTPEPTRLQEGSLSS	1577
Db	1547	TPKAABEKKOINTFVGTPVEK-----LDLIGNP-GS-KRRQYTK-----EKA	1589
Oy	1578	KASOD-----KRLTSTP-----REIA-KSPHSTYVENHHPHISPEYHELNGVSG	1620
Db	1590	KALEDLAGFEKLPFTPGHTESSMTDDKITVECSKSPQ---PDPKPTSSKQKILKISLK	1646
Oy	1621	VDLYRSHIPLA-FDPTSIPRGIPLDAAAAYULPHLAPNFTYRHPYLYRIGYPTAAL	1679
Db	1647	VGVAEEVLPVQKLRQTS-----GKTYQTH-----	1670
Oy	1680	ENRQTIINDYITTSQOMHNHTATAQADMLRG-----SPRESLALNVAAGRGIDL	1734
Db	1671	--RETADG--KSJKAPEKSAKOMLDPAVNTGTERMPRPKEKSAQSLDIAGK--EL	1723
Oy	1735	SOVE-HLPVLVPPGTPTAMDRLAFLTPAFOPSSSRHSSPLSPGAPTHLTPTTSS	1793
Db	1724	FQTDHT-----BESTDDTKTKIACKSP-----PSMDPTPST--	1758
Oy	1794	SERERDRERDRREREKSI--LSTTVEHAP-----IMRPTQSSGSSGSSG	1842
Db	1759	--RRPKYPLGKRDIVEELSLKQLTQTHTDKVPGEDKGINFRETAKOKLDPAVS	1816
Oy	1843	GGGGSSSRPASHAHQHSPISSPTODALOOREVLNITMGKGI-LPAVESKPTVLRS	1901
Db	1817	G-----SKROPTPKAKOPLBDL-AGIKELFORPVCTDKPTHEKT	1857
Oy	1902	STSPVRAATFPPATHCPLGTLDGVPYPTLMEP-----VLLPKEARVA	1946
Db	1858	-TKIACKSPQDPVGT-----PTFKQSKSLRKADVEESLALRKRTSVG	1904
Oy	1947	RP-ERPRADTG-----NAFLAKPPASGLEPASSPSKSEBRPLV-----PVGHA	1992
Db	1905	KAMDPKPAAGDEKQMKAFMGVPOKIDL-EGNLP--GSKWVPQTPKEKADLEDLAGFK	1961
Oy	1993	TIARTPAKH-----LAPHASPPRPAPASAPRHEKQSKPFIQO--LELRS	2040
Db	1962	ELFORPGDCKPTTEKTKIACKSPQDPVDTPASTK--QRPKNLRADVEEFLLR-	2018
Oy	2041	LGHGSSYSPGVEFVSPVSSPLNHDGKLPMHLELDKSHLEGELRPKQGPAPKLGEA	2100
Db	2019	-----KTPSAGAMD-TPKPAVSDEKNITPVT-----PQKIDL	2055
Oy	2101	AHLPHLEPLPSOBSSPPLQTPARVX-----GHQVVTIAOHISEYITD-----Y	2147
Db	2056	GNLPESKXQOPQPKKAEALBDLVGFKLPQTPGHTESSMTDDKITVECSKQPOSESFKT	2115
Oy	2148	TRHNFQOASAPPAFLYSPGASCVULDRPP--SULYLRPPD--HAPARGPHS	2200
Db	2116	SRSSKORLIKPLVX-----VDMKEPFLAVSKLRTSGETTYQTHPEPGDSXI	2163
Oy	2201	BGGKRSPE--PNKTSYLVGGBD-----GLEP-VSPGEMTEPHGSASVAYPLLR	2247
Db	2164	KAFKSSPQOILDPASVSGSRQLTRKEKAEALBDLVDFEKLPSABGHTESM-----	2217
Oy	2248	DEQOTEPSSMGSGDNTSOPAPFSKLTESNSAMVYKSKOEINKLNTNHRNPEYNIS	2307
Db	2218	--TIDKNTKIPCKS-----PELIDTATSTKRCCKTPRKEVKELSAVERLT--QTS	2267
Oy	2308	OPGTEIFNNMPALTGLMTYKSOA-----VOEHNS-----	2337

[illegible]

A:Cross-references: GB:L01475; GB:L01476
A>Note: the codons given for 1438-ASP (GAG) and 1443-Gly (GCA) are inconsistent with the
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL4A4
A:Cross-references: GDB:132673; OMIM:120131
A:Map position: 2q35-2q37
A:Intron: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
A:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidney
C:Superfamily: collagen alpha 1(IV) chain
C:Keyword: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxy
F:1-38/Domain: signal sequence #status predicted <Sig>
F:39-1690/Product: collagen alpha 4(IV) chain #status predicted <Mat>
F:39-61/Domain: amino-terminal nonhelical, NH1 <NH1>
F:62-1466/Region: interrupted helical
F:94-96/Region: cell attachment (R-G-D) motif
F:145-147/Region: cell attachment (R-G-D) motif
F:189-191/Region: cell attachment (R-G-D) motif
F:310-312/Region: cell attachment (R-G-D) motif
F:724-726/Region: cell attachment (R-G-D) motif
F:785-787/Region: cell attachment (R-G-D) motif
F:989-991/Region: cell attachment (R-G-D) motif
F:1212-1214/Region: cell attachment (R-G-D) motif
F:1467-1690/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1471-1569/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1579-1666/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:47-52,55,57,266,400,460,492,494,666,790,828,1099,1131,1294,1317,1375,1407/Disulfide bo
F:142,669/Binding site: carbonylhydrate (Aan) (covalent) #status predicted
F:1480-1566,1513-1569/Disulfide bonds: (or 1480-1569, 1513-1566) #status predicted
F:1525-1531,1634-1641/Disulfide bonds: #status predicted
F:1588-1663,1622-1666/Disulfide bonds: (or 1588-1666, 1622-1663) #status predicted

Query Match 2.6%; Score 339.5; DB 1; Length 1690;
Best Local Similarity 21.1%; Pred. No. 9e-05;
Matches 375; Conservative 107; Mismatches 617; Indels 681; Gaps 92;

QY 761 DTGNGPK-PPATLGADP-----PPGP-----PTPRRTSRAPITPPASEATGATPP 809
DB 380 ETGDVPPPPGGLGPRGECACMTGPPQGFGLPGLRGEGITGRDS-APGRGPR 438
QY 810 PAPPSAPPPVVKKEKEETAAAPVEGEQKPPAAEBELAVDTGK-AEEPVKSECTE 868
DB 439 GSPGLPGAP-----GLQGPGSSVTCVGNPQPGIKGVAP 476
QY 869 EAEERAKGKDAAATAEAGLKAKEKGGSRATTAKSSGAPQSDSSATCSADVD 928
DB 477 PGGRGKGEKNGELCACERPGPPGGLPGRQSKDGLGPMVLTGK----- 526
QY 929 AEGGDNRLLSPRLPTGDRANASPOKPLDLKOLKORAAIIP-----IOY 978
DB 527 ---GD-----PGPGCAGPPGLP-----GKHGSGPRKMGAKGDMVVR 563
QY 979 TKVNEPRDEADATKPAAPAPPONLQPSDAPQPGSSPR---GKSRSPAPADKEA 1034
DB 564 VKGHKGER-----GPDGP-----GPGCGPGSHGRDHAGEKDDPPRPDHD 606
QY 1035 FAAEAKLQCD-----PCWTSGLPFPVPPREVIKASPHADPAFAVAPRGHP---L 1084
DB 607 ATPGGGFPGLPGRGKAGVGPGLGFPGRGE-RGNP-----GVGHHGVNG 654
QY 1085 PLGLH-----DTAPVLPAPPTISNPPLISSAKPSVLEROIQAISQMSVOLAH 1135
DB 655 PDLKQKQKDTISCNVTYPRGRHGRPGFDRP-----GPKGPRGGA--PLSGS--- 702
QY 1136 PYSEHKAAPVGYVTMGLPLPM-----DPKTLAPPSGVKQEQLSPPGQAGPPESIGVPT 1188
DB 703 --DGHKGRPGTPTAETIPGPRGPRGMGDG--GFGGEGKS--SPVGGPRGPGSPGV-- 752

QY 1189 AOEASVLRGTAAGSY--PGSITTKGIPSTRYPSDAITVYRGSITHTGPADVLKGTITRI 1246
DB 753 NGQKIPDDPAFGHLGPRGKGLSGVPIDKGP-----RGD--PGCP----- 791
QY 1247 IGEDSPSRLL-----DRGEDSL-----PKGHVIEGKKG-----HYLSEYSGMSV 1286
DB 792 -GAEPPAGIPGLGLKPKGRGHNAGFPVGPFPHSOSERAPGLPGQGLPGVYG----- 846
QY 1287 TQCKEDRSSSGPHEHTAARITYDMEGRYGALISSASIEGLMGRAIPPERHSPHLK 1346
DB 847 -----SPGAGGKGQF-----GVVGP--GPAGMKGLPGLGRPRHAPPGI-- 886
QY 1347 EGHNRGSITQIPPSVYAEADYLRRAKLLKREGTPEPPPSRDLTEAVYQALGPIK 1406
DB 887 -----PGLPGRPD-----DGLPGRPK-----GRRG 909
QY 1407 L-----KPAHEGLVATVKEAG-RSIEHIEPREELRHTPELPLARPLKEGSIQ 1454
DB 910 LPGFPGRGERKPKAEGCPGAKGEPGKMSGLPGD-----RGLRGAKAIG 957
QY 1455 TPLKYDTGASTGSKKHVRSILGSPGRTPPVHPDLVMADARALERACYESLSKRPCT 1514
DB 958 PP--GDEGEMALISQK-----GTPGPRGP-----GDDGPRGERD 991
QY 1515 ASSSGSIAAGAPVIVPELGKPRQSLTYEDHGAPFAGHLPRGSPVTREPTRLQEGSL 1574
DB 992 KCTPKMGRRG-----ELGR-----YGP--GPHRGEGEKQGPDPGPPGP 1032
QY 1575 SSSKASQPRKLTSTPREIAKSPHSTVPEHNHPISPYEHLRGVSGVDLYRSHIPLADP 1634
DB 1033 GSTGL--RGFIFP-----GLPDQGPGRGPP--PGFSGIDGARG----- 1069
QY 1635 TSIPRGIPLDAAAYLPRHLAPNTYPRHYLYPYLIRGYPTDALENNQTIINDYITQ 1694
DB 1070 ---PKGNKDPAS-----HFGPRG--PKGEPGSPGCG-- 1097
QY 1695 MEHNATATMAQRADMLRGLSPRESSLALNVAGPRGIIDLSQVPHLPVLVPTPTATA 1754
DB 1098 -HFG-----ASEGGLPGI-----GPRG-----SPRPPGPPG-- 1124
QY 1755 MDRLAVLPTAPQPFSSRHSSSPDSPGPTHTLTKPTTSSSERDRDRDREREKSI 1814
DB 1125 ---SSGP--PGCPDHPMGLRGQPGEMD--PGPGLGQDPGI 1161
QY 1815 LNSTTTEHAPLWRPCTEGSSSSSGSGG--GSSSRPASHAHQHSPIPRQDOLQ 1872
DB 1162 -----PPPGIKGPGSPGLNGHLGKQKGTGASGLHD-VGP----- 1199
QY 1873 QRPSVLHTGMKGIITVAVPSKPTVLASTSTSPVR-----AATPPATNCPAGTLDGV 1928
DB 1200 --PGVIGIPLGK--ERQDPSPGI-----SPGPRKQKPPRPPSSSGPPG----- 1242
QY 1929 YPTLMEPVLLPKEARPVARPERPRADTGHAFIAKTPRANSGLPASPSSKGSPPRLVPV 1988
DB 1243 -----PAGATGRAPKDIIPDPGPDQG-----PPGPDGPRGADP-----PGL 1280
QY 1989 SGHATIAATPRAKILARNHASPPRAPASASDPHEKQSKFSLQELTASLGYNG-- 2045
DB 1281 PGSVDLKRPEPDC--GLPAPRPPGPPGP-----GTXKGFPG 1316
QY 2046 -----SSYSPREGEV--PVSFVSSPSLTHDKGLPKHLELDKSHLEGLRPPKQDPVYLGE 2099
DB 1317 CQKQKQKPMKFPREGQPHGFRPGRGKGLG-----PRGRGPTGLPG-- 1361
QY 2100 AALHLRLPESQSSSPL--LOTAPGVKHQVYVTLAQHISEVITDYYTRHNPQLS 2156
DB 1362 ---PRGEGRPADVDPCRIPLPGAGMRPRDEAMGI-----PGR 1400
QY 2157 APLPAPLYFPGASC--PVLDLRPPSDLYLPPRDHGAPARGSPHSEGKRSPPENKTS 2213
DB 1401 GP-----PPGCKGEPGLDGRGVD-----GVP--GSFPGPPKRGDT----- 1435


```

QY 862 V---KSECTEAEAEGBPAKGDAAEAATAEGALKAKEKGGSGRATTAKSGAPODSSS 918
Db 271 VRRRRSRLPKHSDDSGSKHDATATSSSAATVRA-----MSSVGIQTISCS 318
QY 919 ATCSADEVDEAEGBGKNRLSPRPSELTPGTGPRANASPOKRLDLKOLKOPAAAIPPIQV 978
Db 319 VQTEBDQ-----PRVSPA-----IHTNADDPKEI 345
QY 979 TKVHEPPED-----AAPTAPAPAP--PPQNLQPSDAPQOG-----SSPRGKRS 1025
Db 346 VRYSAPKRTGSGESLACOTEPDGOAQGVAGPOLGVPALISPYLGIQIVTGPGLGRREK 405
QY 1026 PAPADKAPAAAB-----AQKLPDPCWTSGLPFPFPPEVIVKASHADPSAFSAP 1079
Db 406 KKPPPLEIGYAAHLPPESLSQVSRQPKSPQVLTSPSP-----LSHRLDLTSPASSE 460
QY 1080 ---RCHPLPLGLHPTA-----RPVLPAP-----PTISNPPILISSAKHPSVLERQIAIS 1126
Db 461 RLNRHAHVSP--OKHFTADSLAKQOTLPFRPMKTLQRLSLSPKPLSPFAEE--SAKER----- 512
QY 1127 QGMSVOLHVPYSEHAKAPVGPVTMGLPLPMDPKKLAPSPGVKQEQSLSPRGOAGPEESLGV 1186
Db 513 --FSLYQH-----QGGLGQVSALP-----PNSLV--RKVXRLTLPSP--PREEHL 552
QY 1187 PTAQAS--VLAGTALGSPGSGITKGIPS--TRVEDSALITRGSLTHGTGADVLYKGTIT 1244
Db 553 PLACQASQOLYAASL-----LQRLGTPPTVPATKA-----SLRELRLDL----- 593
QY 1245 RIIGDPSRLDRGEDSLPKGHVIEGKKGHVLSIEGMSVTVQSKEDGRSSSGPHEP 1304
Db 594 RLVEHES--TKL--RKQALDEBEKEIDAKLTL-----ELGITO----- 630
QY 1305 AAPKRTYDMEGRVGRALISSASIEGLMGRALPPEHSPHMLKEQHHRGINSITQGIPRSIV 1364
Db 631 -----RKESLAKDRG--GRDVPPLGLGEH----- 653
QY 1365 EAQEDYLK--REAKLLKREGTPPPPPSRDLTEAVYTOALGPLKLPALHEGLVATYKEGR 1423
Db 654 ---RDYLSDESLNQRLQSGCTTPAGQFVDF-----PATAAAPAT--PSGP 693
QY 1424 SIHEIPRELHPTPLPAPRLPKESITQGTPLKYDTGASTGSKKHADVRLSLSPGRT 1483
Db 694 TAFQOFRQ-----PPAP-----QYSAGS-----GGRQON 718
QY 1484 FPPVHPLDVMADARALERACYEESLKSRRGTASSGSGIARCAPVIBELGKPRQSPPLY 1543
Db 719 GPPAH-----QAPVFPGETVPAPAPPPGAS--YPAEPGLPNQOAFRP 759
QY 1544 EDHGAFFAGHLPGRS-----PYTWREPTPRLQBSGLSSSKASQDKKLSTPRELKS 1595
Db 760 TGH---YAGQTFMPTQGTLPFPVPASRAPIQKPRQTL---ADLEKVPFNVEVIAISP 812
QY 1596 ---PHSTVEH---HHPHISPYEHLRGVGVDLVRSIHPLAFDTSIPRGIPLDAAAY 1649
Db 813 VVPMSSASSEISYSPAVSSGTE---QGVV-----PEVPRAGDGSISQ----- 853
QY 1650 YLPRHLAENPTYPHLYPYLLRGYPTALENRQTIINDYITSQOMHNT--ATMAORA 1707
Db 854 -----SPAPYF-----SDSHYTSLEQVVPFNMYWIMDISLTXOSTSTADS 896
QY 1708 DMLKGLSRES-----SLALNYAG-----PGIIDLQSVPLP 1741
Db 897 QRLLEPLGSGSGRPEKEPGEVLDGPTLPCCYAGGESESESDSYDPHG-----KGRHLR 951
QY 1742 VLVPEPTPTATAMRLAYLPTAQPFSSRHSSSLPSGCPHTLTKPTTSSSEERPD 1801
Db 952 SM--ESNRPAST-----HYGDSVDYRHGARVKEKYPGEMGPHKPSKLAPALSKREKH 1005
QY 1802 REHDRRERK--SILTSTTVEH--APWRPTEQSSSGSSSS----- 1841
Db 1006 RKQGEQKISKFSFIEAKDVESDLASYPFPVSSSLVSRGRKFODEITTYGLKKNVVEQ 1065

```

```

QY 1842 -----GGGGSSSRPASHS-----HAHQSPIS---PRTOPALQORESVL 1878
Db 1066 KYMGSSRDVAVEDRITGGSSSRAPASAYGCEKLSHDFSGMGYERERAAVNR-----L 1121
QY 1879 HNTGMKGIITAVEBSK---PTVLNSTSTSSPVRAATFPATACPLCGTL----- 1935
Db 1122 QKAGKPSLSLMAHSRVRPMPRSQASEESVSLGRPRA-----GGPLRPGDTCQF 1176
QY 1926 --DCVYPTLMEPVLLPREAPRV---APRERPADTGAFPLAKPRA--RSGLEPASPSKGS 1979
Db 1177 CSHSMBDVOEHV---XQGPRAHAHYKKEEGYILDDSHCVVSDSAHYHLQGETWMDPKPR 1233
QY 1980 EPR--PLVPPVSGHATARTAPKXNLAPRHASDPPEAPASADPREKTSQSPFSTOELT 2038
Db 1234 DARSDRRHNGGHA--VSSSGKGRPAHSTHYDVEPPEBGLWPHDEBGRGHASAKHRH 1292
QY 2039 RSLGHSYSYPGEEVSPVSPSLTHDKLPRLHELDKSHLEGELRPK--OPG----- 2092
Db 1293 GDHRHSGRHT--GEERGRAPAKP---HARDLGHARPHSQPSAPAMPKKGQGVPS 1347
QY 2093 -----PVKLGEAHLPHLRPLPESQPSPL---QTAPEYKGHQV 2132
Db 1348 AEYQPSRASSAYHNASDSKKSRAHSGPALQSKAPEQAPOLQSGROAPG----- 1400
QY 2133 VTLAQHISEVITQDYTHHPQQLSAPLPAPLPSPPGASCPLDLRRPESDLYLPDPH-- 2190
Db 1401 -----PQOSPPSSRQTPS--GAASRPQTQOQOQGLQPPQAL 1439
QY 2191 -----GAPARGSPHSEGKRSPEPNKTSVLGGSDGIEPVSPREGMTEPG 2235
Db 1440 TQARLQOQSOPTRTGSAPASQP--AGKQPGHSTMT-----GQPAEPRAEQONG 1489
QY 2236 HSRSAVTPLYRDEQTEPS-----RMGSKSPGNTSOPPA-----PFSKLTES 2278
Db 1490 SKGTAKAPQOGR--APQAOPTPPGPGAGVAKGAR--PGGTPGAPQAGADESVSKILPG 1547
QY 2279 NSAMVSKKOE 2289
Db 1548 GAEOAGKLTB 1558

```

```

RESULT 69
CGHUB
collagen alpha 4 (IV) chain precursor - human
NAlternate names: procollagen alpha 4 (IV) chain
CSpecies: Homo sapiens (man)
CDate: 06-Feb-1995 #sequence revision 03-Oct-1995 #text_change 16-Jun-2000
CAccession: A55360; S36854; S28777
RLeinonen, A.; Mariyama, M.; Mochizuki, T.; Trygvaeson, K.; Reeder, S.T.
J. Biol. Chem. 269, 26172-26177, 1994
A>Title: Complete primary structure of the human type IV collagen alpha4 (IV) chain. Comp
A|Reference number: A55360; MUID:9501445; PMID:7523402
A|Accession: A55360
A|Status: nucleic acid sequence not shown
A|Molecule type: mRNA
A|Residues: 1-1690 <LEI>
A|Cross-references: GB:X81053; NID:g574805; PIDN:CAA56943.1; PID:g574806
R|Sugimoto, M.; Ohashi, T.; Yoshiooka, H.; Matsuo, N.; Ninomiya, Y.
FEBS Lett. 330, 122-128, 1993
A>Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen cha
A|Reference number: S36854; MUID:93374047; PMID:8365481
A|Accession: S36854
A|Molecule type: DNA; mRNA
A|Residues: 1219-1658, 'FE', 1661-1690 <STG>
A|Cross-references: DDBJ:D1391; NID:g440365; PIDN:BA04214.1; PID:g457161
R|Experimental source: whole eye
R|Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A>Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
A|Reference number: S28777; MUID:93054733; PMID:11429714
A|Accession: S28777
A|Molecule type: DNA
A|Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 <KAM>

```



```

QY 1126 SOGMSVOLHVPESEHAKAPVGVITWGLPLPMDPK-----KLAFFSGVK-QEOL 1172
DB 3013 PKLQSVQ-----EKSQVPRVVK-----PLDPKPOKEQESKNEKNEKVPSTNVLDDQA 3060
QY 1173 SPRGAGPESIGVPTAQASVLRG-----TALGSPGSGITKGIISTVSPDSALTVYGS 1228
DB 3061 SSSNQGGGSAKPSASIFTLPRKKEKMTSTASSSIPISPENQEAETSTNKS 3120
QY 1229 IHGPRADVLVNGITIRIIGEDSPSRDLGRDSTLRKGVYEGKKGHVLSYEGMSVTO 1288
DB 3121 -----PIERIKKERTGOLIKRSPSDPKPCV-----SGALVFKVDKPI-S 3163
QY 1289 CSKEDRSSGPPHETPAKRTYDMWEGVRAISSASIEGLMGRALPPRHSHPHLKQ 1348
DB 3164 -----NSPYGKPMARRPKA-----DSAADSDLG----- 3190
QY 1349 HHIRSGITGIPRSYAEODVIRRAKLLKREGTPPPPPRDLTAAYTQALPLK 1408
DB 3191 -----SDEGSTSHGCLKDENSQKP-----AGKTVRFGPISTY 3225
QY 1409 PAHEGLVATVKEAGRSIHIEPRELHTPEL-----PLAPRPLKE-----GSITQGTPLK 1458
DB 3226 PEPAPAPQV-----ELPKVE-----PTPEISFESLAQQAQKTLAKEASKILREYVPT 3275
QY 1459 YDTGAST-----TGSKKHVRSLLIG-SPGRTPPVHPPLVMDARA-----LERACTEE 1506
DB 3276 EGTSPETAEPVSPTKSGSK-----LSFFGLSP-----FGKVTPEKNTAQTOSTLVLESTSP 3328
QY 1507 SL-KSRPGTASSSGSIA-----RCAPVIVPELQPRQ-----SP 1540
DB 3329 LTPSPNETSNKSTSSKOPKSLSEKTIATOMIQTRGP-----IKLSHRQOLKIDPSP 3386
QY 1541 LTYEDGAPAGHLP-R-GSPVTKEPTPRLQESLSSKASQDKLTSTPREIASPHST 1599
DB 3387 LQSGSS-----LPRAPQSPKPPQLEVPQGVQVTPNPSQKRYTRLGRREKPPPKAK 3441
QY 1600 VPEHHPIPIPYEHLRGVGVLYRSHITPLADPPIISPIGILDAAYIYP----- 1652
DB 3442 TPLSIRVPOS-----GVS-----ISTP-AIDLAKVELIADIEKTKM 3478
QY 1653 RHLAPNTPHLYPVYIRGVPPTAL--ENROTIINDYITSQO----- 1694
DB 3479 RKRKPKMETTL-----VLDGTEHFKULTAENIQLSOPKSKKMKKRTKWSVIGCB 3534
QY 1695 -----MHNTATAMAQRADMLRGSPRESSALNTAAGPRGIIIDLSQVPHLPVLPPI 1747
DB 3535 NDSKTCMOHN--KLLEFRDLKSLITRQ-----LDEKMI----- 3568
QY 1748 PGTPATAMDRLAYLPPTAPOPSSSHSSPLSPGPTHLTKPTTSSSERDRDRER-DR 1806
DB 3569 -----TAPAVFNPRP-----TSSSTKHLLNKKPKFEY 3597
QY 1807 DREKESILNTSTTVEHAPIWRPGTEQSSGSGSGSGSSSRPASHAHQSPISPR 1866
DB 3598 TRKKLKLANN-----RQCKIRMSG-----IHSQPKDVAISAI 3631
QY 1867 TODALOQPSVYANTGKGIITAVEPKPYLRSTSSPYRPAATPPATHCPISGTL 1926
DB 3632 RQ--VQKVFQSTGKKMNPILISGFRP-----KSLIKPEPKPATV-----LFRIDEGN 3679
QY 1927 GVPYTLMEPVLLPEAPRVARPER--PRADTGHAPL--AKPARAGLEPASPSGSE 1980
DB 3680 KLTKKLGSPDMQKNDKCNPMASMPBRKKSNSYLKRNAMAKWAKG--KKMWTLFE 3736
QY 1981 PRPIVPPVSGHATTARPPAKNLAPHHASDP--APPASADPHREKTOGKPSIOLE 2037
DB 3737 TNOJLTRLG-----APKDSHPKDVLVDPKADAPPSPRS--LSWTTVSSNALSNSL- 3786
QY 2038 LRSIGYHSSSYSPBEVPEVSPSLTHDKGLPKHLELDKSHLEGLRP-----KOP 2091
DB 3787 LDALNTVGS--EPSTSSGIVTDIAPVKEQFRKAK--ESRPTQOTVSKKY 3834

```

```

QY 2092 GPVKLGEAHLPHLRPLPSPSSPPLQOTAPGVKHQGVVTLAQHISEVI----- 2143
DB 3835 PYYTLKNTVLR-----SPSSSTSL--VNLDEREAVSSLAQDGSIIYEVGQEEB 3881
QY 2144 TQDYTRHHPOOLASAPLAPLYSFPGASCPV--LDLRPPSDLYLPPDHGAPARSP-H 2199
DB 3882 VETTKRPDPFVAPLISALFKAFPTSSSEFVGLBADSRSRTERTL-----RGLFRSSPTM 3937
QY 2200 SEGGR--SEPKNTSVLGGGEGDIEVSPPEGTEPCHRSAAVYPLLYR--DGEOTEPS 2255
DB 3938 PEFSTRVSSSPTRRRNRG-----PVTSSQSSAEDILALMQAPDALENMEE 3987
QY 2256 RMGSKSPGNTSQPPAFPSK-----LTESNAAWYKSKOINKKLTNNRNEP 2302
DB 3988 QVSKFPPIQALVAVLNGKGMWRPKLKNLSVTOFRELKESSKRXLYEVLQNYKNK 4047
QY 2303 EYNISQPTIEFNPAITGLMTYRSQAOVHAETMGLAIIKRLMKRYQWESSP 2362
DB 4048 SYSTLQT-----LITFN--LDQYKRWLYRPPKLTMDLQKIDP----- 4084
QY 2363 LSANAPNPNMAS 2376
DB 4085 -----NSLRVSS 4092

```

RESULT 66

A6194
neurofilament protein NF-220, high-molecular-weight splice form - longfin squid
C1:Species: Loligo pealeii (longfin squid)
C1:Date: 22-Sep-1993 #sequence revision 25-Apr-1997 #text_change 17-Mar-2000
C1:Accession: A6194
R1:Way, J.; Hellmich, M.R.; Jaffe, H.; Szaro, B.; Pant, H.C.; Garner, H.; Batey, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 6963-6967, 1992
A1:Title: A high-molecular-weight squid neurofilament protein contains a lamin-like rod d
A1:Reference number: A6194; MUID:92357751; PMID:1379729
A1:Accession: A6194
A1:Status: nucleic acid sequence not shown; not compared with conceptual translation
A1:Molecule type: mRNA
A1:Residues: 1-1200 <MAV>
A1:Cross-references: GB:M94389; NID:g161291; PID:g161292
A1:Experimental source: stellate ganglion
A1:Note: sequence extracted from NCBI backbone (NCBI:113499)
C1:Superfamily: neurofilament triplet H protein
C1:Keywords: alternative splicing

Query Match 2.6%; Score 341; DB 2; Length 1200;

Best Local Similarity 19.4%; Pred. No. 5.5e-05;

Matches 237; Conservative 196; Mismatches 491; Indels 298; Gaps 47;

```

QY 137 KDRSLTGKLEVPSPSPHTDELEVPRLSKELIQMDRVRE-ITM---VEQGISK 192
DB 120 QMKKLAGELELEKSKWKGETSAIKEMYETEL--EBAKLIATNKEKITLQVRYTELIDQ 177
QY 193 LKKKQOQLEEEAAKPEPEKPVSPPIESKHSLVQIYDENRKAFAAHLRLEGIAPQV 252
DB 178 LERQQDLDES-----RTYQIDDEQIARQNGQLAD-----LEG----- 211
QY 253 ELPLYNQPSDT-----ROYHENIKINQAMERKLI-----LYF 284
DB 212 EISMRLRSIESLEKEMQSNILAQNDEMCKRMDLNNETINHLDAENRRQTLDEELF 271
QY 285 KRANHA-----RKQKQKCCQRYDQMLELLEKVRIRIENPPRRAKE 326
DB 272 QKVVAHQELKEALALAYRDTTAENEFWNEIQAIRDIQOEYDAKCDQMRGD----- 324
QY 327 SKVREYYE-----KOPPEIRKQRELEQRMQS--RVGQSGSLSMSAASEHEV 372
DB 325 --TEAVYNLKVQEFRTGATKQMEVYRNKEENTYKLSMNTETIRNLADLEAANAQLEERN 382
QY 373 SEIIDGISEQENLEKQROLAVIPMLYDADQRIKFINMGLADPMKAYVYDROQVMNM 432
DB 383 QDLRLDLEKDRQNE-----LESCQYKEEITYLRGEMESILKEIQLDMLDIKLS 430

```



```

Db      376 HGAQGP---RPPGNGSPG-----GKGMGPAGIGAPGLMGARGPPGP 418
Qy      1148 V-TWGLP-----LPMDPKLAPSGYKQOLSPGQAGPPESIGVPTAGASVLRGT---- 1198
Db      449 AGANAPGRLRGAGGEGK-----NGAKGEP-GPREGGEGAGIPGVPAKGEDEKDGSPGP 473
Qy      1199 AAGSGGSGITGISTRTVPSDAITRGSIITHGPRADVLVYKGTITRIIGEDSPRLSD 1258
Db      474 GANGLPGAAAGERGAPGFRGPAG-----PAGLP-----GEGKPGAG-ERG 510
Qy      1259 RBDLSL-PKGNVLYEGKKGHVLYEGGMSVYQCSKEDGRSS-----GPPIHTAAPKATYDM 1313
Db      511 APGAPGPRGAPGAPGPDG-----VPGGPRKMGMPGSGPGRGDKKPPPS----- 556
Qy      1314 MEGVRG-----AISASIEGLMGRAITPERRH-SPHILKEQHINRISITOGIPRSTVEAQ 1367
Db      557 -OGESGRPPGPGSGRGPQGVWGFPGKNDGAPKNGERGPGGPGQGGPKNGKENG 615
Qy      1368 EDYLREAKTLKREGPPPPPPSRDLTEAYKTQALGRLKKAHEGLVATVEAGRSIH 1427
Db      616 P-----QGPPTGPGDKGDT-----GP-----PDPQGLQGLPTGCP----- 649
Qy      1428 IPREELRHTPELPLAPRLKESITQGTPLKYDTGASTTGSKKHVRSILGSPGRTFPPV 1487
Db      650 -PGENCK-----PGEPRGP-----KGDAGAPGAPGKGDA-----GAPGERGP- 686
Qy      1488 HPLDVMADRALERACYEESLSKRPCTASSSGSLAKGAPVLYPELCKRQSPLYTEDHG 1547
Db      687 -----GLAGAPGL-----RGA--GPP--GPEGKGAAGP-----PG 714
Qy      1548 APFAGHLPRGSPVYMEPTPLQEGSLSSSKASQDKLNTSTPEILAKSPHSTVPEHHP 1607
Db      715 PEGAA-----GTRELQMPGER--GGLSGPGRGDKGEGEGRP----- 750
Qy      1608 ISPVEHLRGVSGVDLYRSHIPLAFDPTSI PRGIPLDAAYLYLPRHLAENPTYPLYP 1667
Db      751 -----ADGVPGKD-----GPRG-----PFGP-IGPP 770
Qy      1668 YLIRGYPDTAALERNQTIINDYITSQOMHNATATAMAQADMLRGISPRESSIALNYAAG 1727
Db      771 -----GPAQGRQDKKEGAP-----GLPGIAG 792
Qy      1728 PRGIIIDSGVPHLVLPPTPGTPA--TAMDRLAYLTPA-----QPRSSHSSSP--LSQG 1780
Db      793 PFG-----SPGERGTGPPRAGFGAGGQNGRPGKGERGAPGKRG 836
Qy      1781 GPTHLTKPTTSSERRDRDRDRERERKSILNTTVEHAPIMRGTBOSSGSSGS 1840
Db      837 GPPGVAGPPGGS-----GPAQPPGQGVKGERGS 865
Qy      1841 SGGGG-----GSSSRPASHAHQHSPISPRTODALQORPSVLANTGMGIITAVPSKP 1895
Db      866 PGQPGAGPFGARGLGPPGSGNNGP PGPSPSGPGKGPFGAGNTGARG- 915
Qy      1896 TYLRSTSTSPV-----RPAATFPATHCPLGTLIDGVYPTLMEPVLLPKKAPRY 1945
Db      916 ----SPGVGPKGDAGQPGKSGPQAQGPARGPLG--IAGI----- 952
Qy      1946 APERPRADTGAFLAKPPRARGLEBPASSPS-----KGSBRPLVPPVSGHATITARTAKN 2001
Db      953 -----TGARGLAGPPGMPG--PRGSPGQGVKGESGKP-----GANG 987
Qy      2002 LAPHNASPP--PAPPASADPHREKTQSKPSIOLELRSLGHHGSSYSPEGV----E 2054
Db      988 LSGERGPGRPGGLPGLAGTAGBERGDNPG-----SDGLRGDGSPPGKDKRGE 1036
Qy      2055 PVSVPSPSLTHDKGLPKHLELDKSHLEGELRPKQPGPVKLGEAAHNPHLRPLTESOP 2114
Db      1037 NGSP-GAPRAGHGPGRPGVPGAKSGDGESGP--AGPAGAPRAGS-----RGAGPFG 1089
Qy      2115 SSSPLLQT---APGVGHQORVVTLLAQHTISEVTITDYTEHNPQOLSAP-LPAD-----LY 2164
Db      1090 PRDCKETGERGAAGIKGH-----RGFPGNPAGSPGAPGQQA 1130

```

```

Qy      2165 SRFGASCPVLDLRP-----PSDLYLPPPDHGA PARSGPHSGKRSPP 2209
Db      1111 GSPGAPG---RGVSGSGPPGKDGTSCHPEIPPPGGRNGRSGSGSPGPGQPG 1186
Qy      2210 -----NKTSLVG-GGED-----GIEP-----VSPPEGTEPGHSRASY 2242
Db      1187 PGPAPGPGCCGVGAALAGIGEKAGGAPAYYDEPMDFKINDELMTLSKANQGLE 1246
Qy      2243 PLYRDEGQTEPSR 2256
Db      1247 SLSPDGSRKNDP 1260

RESULT 64
OFHGH
neurofilament triplet H protein - human
N:Alternate names: neurofilament protein, 112K
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
C:Accession: S00979
R:Lees, J.F.; Shneidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
EMBO J. 7, 1947-1955, 1988
A:Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
A:Reference number: S00979; MUID:88328981; PMID:3138108
A:Accession: S00979
A:Molecule type: DNA
A:Residues: 1-1020 <LEB>
A:Cross-References: EMBL:X15306; NID:g35028; PID:CAA33366.1; PID:g1841430
A>Note: It is uncertain whether Met-1 or Met-2 is the initiator
C:Genetics:
A:Gene: GDB:NEFH
A:Cross-References: GDB:120225; OMIM:162230
A:Map position: 22q12.1-22q13.1
A:Introns: 295/L; 361/3; 403/2
C:Superfamily: neurofilament triplet H protein
C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
F:1-100/Domain: amino-terminal <NTD>
F:101-410/Domain: rod #status predicted <ROD>
F:411-1020/Domain: carboxyl-terminal <CTD>
F:502-826/Region: 14-residue repeats
F:503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,644
(covalent) #status predicted
F:732,768/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match      2.6%; Score 342.5; DB 1; Length 1020;
Best Local Similarity 20.6%; Pred. No. 4,1e-05;
Matches 201; Conservative 147; Mismatches 398; Indels 231; Gaps 32;

Qy      166 RLSKEELIQMDRVREITMVQISKLKKQOOLEEBAKPPPEKPVSPPIESKHS 225
Db      168 RLQEHLELDIAHV-----RQRDDA----- 189
Qy      226 LVQIITYDENRKAQAARHL-----EGIGPOVEL-----PLYNPSDTRQYHE----- 268
Db      190 -----RQREBAEAARALARFAQEAARAVLDQKQAOLQEGCYLRHHQEEVGEEL 242
Qy      269 -----NIKINQMRKKLILYFRKRNARK---QWRKQFCQRYDQ 304
Db      243 GOIQSGAAQAOQAEETRDALKCDVTSALR---TRAQLEGHAVOSTQSEEFVFRLLDR 299
Qy      305 LMEALEKKVERIENNPBRRAKESKVREYERQPEIRKQRELOERNQSHVGQSGLSJS 364
Db      300 LSEBAK-----VNTDARKSAQEBITTEY-----RROQATTELEALKSTYDSJE 343
Qy      365 AARSEHVESEIIDGSEQENLEKQKROLAVIPMLVDADQRIKFINMGMLADPMTKYK 424
Db      344 RQRESELEDRHQADIASYQGAIOQ-----LDAELRNTKM-----EWAADLREYO 386
Qy      425 DRQVNMNMSEQEKETPREKFMQHPK---NFGILASFLEKTYAECLVLYYLTKXNENYKS 481
Db      387 DILNVMALDITIAVYR-KLLEGECRIGFGPIPSLPGRLKIPSVSTHIVKSEBKIK 445

```

A>Title: G to T transversion at position +5 of a splice donor site causes skipping of th

A:Reference number: 155349; MUID:91161621; PMID:1672129

A:Accession: 155349

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 537-605 <LE>

A:Cross-references: GB:M59312; NID:g180815; PIDN:AAA52041.1; PID:g180816

R:Sever, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980

A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty

A:Reference number: A90438; MUID:80198282; PMID:6246925

A:Accession: A90438

A:Molecule type: protein

A:Residues: 728-895, 'A', 897-964 <SEY4>

A:Experimental source: liver

R:Cole, W.G.; Chido, A.A.; Iamande, S.R.; Janeczek, R.; Ramirez, F.; Dahl, H.H.M.; Chan

J. Biol. Chem. 265, 17070-17077, 1990

A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an

A:Reference number: A38303; MUID:91009133; PMID:2145268

A:Accession: A38303

A:Molecule type: mRNA

A:Residues: 861-1015 <COL>

A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:g180878; PIDN:AA59383.1; PID:g

A>Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos sy

R:Mancko, B.S.; Dalgleish, R.

Nucleic Acids Res. 16, 2337, 1988

A>Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.

A:Reference number: S02119; MUID:88189827; PMID:3357782

A:Accession: S02119

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>

A:Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA2986.1; PID:g30054

R:Sever, J.M.; Kang, A.H.

Biochemistry 20, 2621-2627, 1981

A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty

A:Reference number: A90446; MUID:81208139; PMID:7016180

A:Accession: A90446

A:Molecule type: protein

A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-

A:Experimental source: liver

R:Lofti, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Wye

Nucleic Acids Res. 12, 9383-9394, 1984

A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage

A:Reference number: A93551; MUID:85087944; PMID:6096827

A:Accession: A93551

A:Molecule type: mRNA

A:Residues: 1065-1155, 'P', 1157-1466 <LOI>

A:Cross-references: EMBL:X01655; EMBL:X01742; NID:g29584; PIDN:CAA25821.1

R:Miskulini, M.; Dalgleish, R.; Kluge-Beckerman, B.; Renard, S.I.; Tolstoshev, P.; Brann

Biochemistry 25, 1408-1413, 1986

A>Title: Human type III collagen gene expression is coordinately modulated with the type

A:Reference number: 152393; MUID:86187804; PMID:3754462

A:Accession: 152393

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1161-1200 <MIS>

A:Cross-references: GB:M13146; NID:g180415; PIDN:AAA52003.1; PID:g180416

R:Emanuel, B.S.; Camilizaro, L.A.; Seyer, J.M.; Myers, J.C.

Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985

A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm

A:Reference number: 159025; MUID:85216505; PMID:3858826

A:Accession: 179359

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1165-1196 <ENWA>

A:Cross-references: GB:M1114; NID:g180417; PIDN:AAA52004.1; PID:g180418

R:Chu, M.L.; Well, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.

J. Biol. Chem. 260, 4357-4363, 1985

A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen.

A:Reference number: A92516; MUID:85157600; PMID:2579949

A:Accession: A92516

A:Molecule type: DNA


```

QY 1719 STALVYAGPRGIIIDLSQVPHLPVLVPTPTATAMRLAVLPTAPQPFSSRHSSPLS 1778
      |||||
      ---GLGIAGPRG-----GPG-----EGEGHPREPAPF----- 810
Db 785
QY 1779 PGGPHLTKPTTSSERERDRDREREKSLITTTVEHAPLWRPQTEOSSGSS 1838
      |||||
      ---CAKGERGAPGEGGPPGAPGPGTSS-----GPAPPGPGQVKGGR 862
Db 811 PGAPCGNGP
QY 1839 GSSG-----GGGSSSRPASHAHGHSPISTPTDALQORPSVLNMTMKGITV 1890
      |||||
      GSPGPPGTAGPPGGGGLPBPNNNGP---GPPGSGADKDGKPPAGNSG----- 911
Db 863
QY 1891 ESKPTVLRTSTSTSPV---RPAATFPATHCPLGTIDGVYPTLMEVLLPKAPRVA 1946
      |||||
      912 SPNGGAGPKDADQPGKGPFGAGGPPGSPGLG---IAGL----- 951
Db 912
QY 1947 RPERPADTGAHFLAKPPASGLSPASSPS---KSEPRPLVPVSGHATTARTPAKNTL 2002
      |||||
      952 -----TGARGLAGPPGMPG---PRGSPGQIKGESGKP-----GA 984
Db 952
QY 2003 APHHASPDPRAP-----PASADPHRETKQSKPSIQLELRSLCYHSSSTSPGVEPV 2056
      |||||
      985 SGHNEERGPQGGLPQGPGTAGPGRDGNPG-----SDGQPERDGSFGKGDPR 1033
Db 985
QY 2057 SPVSSPSLTHDKGLPKHLELDKSHLEGELRPKQGPVYKLGGAHMLPHLRPLPSOPSS 2116
      |||||
      1034 GENSGFGA---FGAGH-----FGPPEPVPSGSKSGDRGTGP---AGSGG 1073
Db 1034
QY 2117 SP---LLOTPA-----GVKQHORVTLAHISEVITQDYTRHPQOL 2155
      |||||
      1074 APGAPGARGAPGQPGRGDKETGERSNGIKGR----- 1108
Db 1074
QY 2156 SAPLPAPLXFPGACPVLDLRPPSDLYLPPBDGAP---ARSGHSSEGGKSPERNK 2211
      |||||
      1109 -----GPPG-----NPG---PRGSPGAGHOGAIGSPGAPGRGVPVGPBG 1145
Db 1109
QY 2212 TSVLGGEGDLE---PVSP-----EGMTEPCHSHSAVYPLLRYRDEQTEPSR 2256
      |||||
      1146 PP---GKDGTSGHGPICPFPGRNNGRGRSGE---SPGH-----FGQGPPEP 1188
Db 1146
QY 2257 MGSKSP-----GNTS-----QPPAFSKLTESNANMYSKKQKQKINKLN 2295
      |||||
      1189 PGAPPCGCGGAAIAGVGEGSGGFPYGGDDPMDFKINTBEINSLSKSVNGQIESLIS 1248
Db 1189
QY 2296 TH-NNEPEYN 2305
      |||||
      1249 PDGSRKNPARN 1259
Db 1249

```

RESULT 63

```

CGHUTL
collagen alpha 1(III) chain precursor - human
N/Alternate names: procollagen alpha 1(III) chain
C/Species: Homo sapiens (man)
C/Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000
C/Accession: S05272; S04642; E0011; S01726; S04687; A90399; A94562; I51868; S59511; A90
R/Prockop, D.J.
submitted to the EMBL Data Library, February 1989
A/Reference number: S05272
A/Accession: S05272
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1240, 'V', 1242-1466 <PRC>
A/Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuitvanleml, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A/Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of huma
erences.
A/Reference number: S04642; MUID:89350838; PMID:2764886
A/Accession: S04642
A/Molecule type: mRNA
A/Residues: 1-1196 <ALA>

```

```

A/Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
A/Note: the complete sequence is not shown
R:Benson-Chanda, V.; Su, M.W.; Weill, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A/Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (
A/Reference number: E0011; MUID:89378752; PMID:2777083
A/Accession: E0011
A/Molecule type: DNA
A/Residues: 1-176 <BN>
A/Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814
R:Itoman, P.D.; Ricca, G.A.; de Crombrughe, B.
Nucleic Acids Res. 16, 7201, 1988
A/Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A/Reference number: S01726; MUID:88303360; PMID:3405773
A/Accession: S01726
A/Molecule type: mRNA
A/Residues: 1-170 <TM>
A/Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061
A/Note: the authors translated the codon CAG for residue 154 as His
R:Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A/Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen
A/Reference number: S04887; MUID:89386015; PMID:2780304
A/Accession: S04887
A/Molecule type: mRNA
A/Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A/Cross-references: EMBL:X15332; NID:g29545; PIDN:CAA33387.1; PID:g30045
A/Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R:Sever, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A/Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A/Reference number: A90399; MUID:77134724; PMID:557335
A/Accession: A90399
A/Molecule type: protein
A/Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A/Experimental source: liver
A/Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
R:Sever, J.M.
submitted to the Atlas, December 1977
A/Reference number: A94562
A/Accession: A94562
A/Molecule type: protein
A/Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A/Experimental source: liver
A/Note: author submitted corrections to A90399
R:Miliewicz, D.M.; Wiltz, A.W.; Smith, A.C.; Manchester, D.K.; Walstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A/Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
fepring.
A/Reference number: I51868; MUID:93304430; PMID:8317500
A/Accession: I51868
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 186-194 <MIL>
A/Cross-references: GB:S62925; NID:g386425; PIDN:AD1337.1; PID:g4261637
R:Chodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A/Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3
A/Reference number: S59511; MUID:96067614; PMID:7487954
A/Accession: S59511
A/Molecule type: mRNA
A/Residues: 302-423 <CHI>
A/Cross-references: GB:S79877; NID:g1195576; PIDN:AA835615.1; PID:g1195577
R:Sever, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A/Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe
A/Reference number: A90414; MUID:79000343; PMID:687591
A/Accession: A90414
A/Molecule type: protein
A/Residues: 399-675, 'N', 677-727 <SEY3>
A/Experimental source: liver
R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991

```



```

Db 161 ESCEKLECEGGERPQOOPHRAORS PQOOPRLHRPQOPEQODPTPEEETLE 220
Qy 720 EA-----EALHAGNEVPBECGPATVNNSTHESI PSHTEAKOTGONGPK---P 769
Db 221 SSLPLPLEAAH--GPRGLKGBKGPVAVLEPOMLEGGPPG-----EGPAGLIGP 268
Qy 770 PATIGADGP--PPGPPTPRRTSRAPRIETPASEATGAPTPPAPPSAP-----818
Db 269 PGIQNPBPVADPGERPPGAG-----LPQSD--GAPPGT--SLMLPRFGSGGCD 318
Qy 819 -PVPVPEKEKEE-----TAAAPVBESEBQKPPAEELAVDTGAEP---VXSE 865
Db 319 KGPVVAAGQAQAILQOARLALRGP PGPWGTGSP-----GLQGPBGSGGLKE 368
Qy 866 CTEAEAEER-----AKGDAEAEATAGALKAEKKGSGGATTAKSAGAPQSDS 917
Db 369 SGDLQOQPRPGOGLTGLKAGRGRAGPDCARLTLD PCKVGRGDFGLPGLRGEKXH 428
Qy 918 SATCSADSV-----DEAGGDKNRLSPRSLTPTGDPANASPOKPLDKQLKQRAAA 972
Db 429 RGDTPRGLPGRPGEDGERDDGEI--GPRGLPESGPKLIGPKP-----473
Qy 973 IPIQVTKVHEPRREDAATKAPAPAPPPOMLQESDAPQPGSS--PRKSRSPAPPA 1030
Db 474 -----PGIRPGPGVGM-----DGPQPKSLGPGGB--PGPBG 505
Qy 1031 DKEAPAAOKLPDPPCWTSGLPFPVPREVITKASPHAPDSASVAPRGHPLPLGLHD 1090
Db 506 Q-----OGTPG-----TOGLPGR-----OGAIPHGKKGQO---532
Qy 1091 TAPVLPAPPTISNPPLISSAKHPSVLEROI GALSQGSVQLAHVSEHAKAPVPTM 1150
Db 533 --KGLPMPGSDGP-----GHPG-----XKGP--PGTK 558
Qy 1151 GLPLPMDPKLAPFSGVQEOQLSPRGQAPRESLGVPTAOEASVLRGTALGSPGSGITX 1210
Db 559 GKP-----GPGQOGLPGVPGPQGVKGVNG--IRGLKHGKEX 594
Qy 1211 GIPSTRVSDAITYRGS1-THTGTPADVLYKGTTRIIIGEDSPSLDGRSDLPKGVIT 1269
Db 595 G-----EDGPPGKGDIGVKGDRGEVGPSS--RGEDEPEG-----PKGRITG 634
Qy 1270 YEGKKG--HVLSEGGMSVTOQSKEDGRSSGPRHETAPKRTYOMMEGRVRAISASI 1327
Db 635 PTGDGPFGMLGEEKLGVPLGIPYGR--QCPKKS LGP-----GPRC-ASGEKGA 683
Qy 1328 EGLMGRAPRERHSPHILKEOHHTIRGS1-----TOGIPRSYVEAOE 1368
Db 684 RGLSGSKGPRGGRPHSGTGQRGPRGATGKSGAKTSGDDPHGPRGERGLPGR-----737
Qy 1369 DYLRREAKLLKREGT--PPPPSRDITLTAAYKTQ-----ALGPKLKPAHEGLVA 1416
Db 738 -----SGSNGPFGKGPGRPGKGLPCHPQORSEVGFQKKGP--PGPVGAVG 784
Qy 1417 TVKEAGRSIHILPRELHTHTPLPLAPPLKESG1TGGTPLYUDTGAITSKKHIDVSL 1476
Db 785 POGAAGET--GPMSEBCH--PPPPRPG--QGLP-----GTAG--KGTGKD 824
Qy 1477 TGSPP-----RTPPVAPLDMADARALERACYEESLKSHPGTAASSG--GS1AR 1524
Db 825 PGPFGAPGKDPAGLRGFPGERGLPGTAGGPKLG--NEGSGPPGAPGSGERGEAGS 881
Qy 1525 GAPVIVBELGKRGQPLTYEDHGAPFAGHLPRGSPVVTREPRPLQOESLSSKASQDRK 1584
Db 882 GGP--IGRGPRG-----PGPPGAAGEKVPBKEKGTG-----913
Qy 1585 LTSPREIAKSPHSTVPEHNPRIPEYELHAGVGVLDYRSHIPLAFDPTSIPIGIPLD 1644
Db 914 --PTGRDVGQPGVL-----PGPAPR--PGVAGEBQDKGV-----GVREQ 950
Qy 1645 AAAAYLPRHLAPNDTYPHLYRPGYPTAALENQTLINDYITSQOMHNTATAMA 1704
Db 951 KGTGKNKEHGPGRGP-IGPV-----GPGAAAGDGAAGL-----GTQGHFG-----A 995

```

```

Qy 1705 QRADMLGELSPRESSLALNTVAGPRGIIIDLSQVHLPVLVPTPCTPATAMDBLALYPTA 1764
Db 996 KADGTGFMNPPPPPTIGLIGLPSPG--EKETEDVGMGPBG-----PGP 1040
Qy 1765 PGPSSHHSSSPLP-----GGPTLTKPTTSSSRERDRDREREREKSIYTS 1817
Db 1041 RGPAGPNCADOPQGPFGVGNLGRFGEKGEBSGSGIQQEPVKGRGRGKGG--ES 1096
Qy 1818 TTYEHAPVWRPGTEQSSGSSGSGGGG--GSSSRPASHAHQHSPIPTOTALQORP 1875
Db 1099 GQPEBGP--PGAKGQGDGPKGNPGVFPDPDPGPGEG--GPRGODGAK---1146
Qy 1876 SVLHNTGKGIITVVEPSKPTVLRSTSSSVVRAATFPRTTHCPGLGGTLGVPTLMEP 1935
Db 1147 --GDRGEDG-----EPQRP-----GSPGTGNGPPG--PLG-----1174
Qy 1936 VILPKEAPRVARPERPRADTGAFLAKRPASGLEPASPEKSGEPRLVPVSGHATIA 1995
Db 1175 -----KRG--PAGSP--GSEGR-----QCGKGA 1194
Qy 1996 RTPAKNTLAPHASDPDPAPASADPHREKTQSKPFI--QELRLSLGTHSSYSPEVE 2054
Db 1195 GDPALIGAPKCTGCVGPAFGKPGP--DGLRGLPFSVGOQGRPGATGOAG---PG--1246
Qy 2055 PVSFVSSPSLTHDKGLPRHLELDKSH--LEGELRPKQPGVYKLGGEAAHLPHLRPES 2112
Db 1247 FVGPPGLRGLGDG--AKGEKHPLGLGLP--PGQGEKD-----RGLBP 1292
Qy 2113 QPSSPPLQTPAGVGNORVVTTLAQHISEVITODYTHHPOOLSAPLAPLYSPGASCP 2172
Db 1293 QGS-----PGKXGEMGI-----PGASGP 1310
Qy 2173 VLDLRPPSDLYLRPPDHGAPARSGPHSBEKGRPEBNKTSVLOGGEDGIPVSPBGMT 2232
Db 1311 I-----GPGG--PP--GLGPAGPKAKATGP-----GPRGKGVGQPPG--1348
Qy 2233 EPGHSRAVYPLVRYDQTEPRBMGSK-----SPGNTSQPPAFPSKL 2275
Db 1349 HPGPGEVITQPLPIQMPKTRRSVDGSKLMQEDBALPTGGAPSGGLIEIFGSL 1403

RESULT 61
149505
adenomatous polyposis coli protein - mouse
N:Alternate names: APC
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: 149505
R:Su. L.
Science 256, 668-670, 1992
A>Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the
A:Reference number: 149505; MUID:92263101; PMID:1350108
A:Accession: 149505
A:Stetus: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2845 <RES>
A:Cross-references: GB:M8127; NID:g191991; PIDN:AMB59632.1; PID:g191992
C:Superfamily: adenomatous polyposis coli protein

Query Match 2.6%; Score 344; DB 2; Length 2845;
Best local similarity 19.0%; Pred. No. 0.00011;
Matches 516; Conservative 334; Mismatches 1073; Indels 798; Gaps 124;

Qy 37 VGLLEYQHSRDVASHLSPGSIIOPRRRPSLSEFQGNRSQEL-----HL 84
Db 615 VGLTYRQNTWTLAIEGGGIL--RNVSLLIAT--NBDHRQILENNCLQTLLOHL 667
Qy 85 R-----PESHYLPBLKSEM--ETESKRRLLELPDLRPS 121
Db 668 KSHSLITVSNAAGTLMNISARNPKQOELMDMGVSMKLNLIHSHK-----713
Qy 122 PLATGQAGSEDLTKORSLTGKLEVPSPRPPTDELLE---VPRLSKEELLQNM 176

```

Db	794	-----	PGRRGKG-----	ADGIRLCKTKGE-----	814		
Qy	1614	LKRGVGVLDYRSHIPLAFDPTSI-----	PRGI-----	PLDAAAYYLPHILAPPTPHL	1664		
Db	815	--KSEDFPRGKGMGIKGRGELGPRGPRGEDPRGKRGKGNDRPGRLGPRGEKGL			872		
Qy	1665	YPRVLINGPYTALENKQTTIINLYISQCHNHTATAMAGRAMDLKGLSPRESSLANY			1724		
Db	873	GVPEL-PGYL-----	GRQ-----		884		
Qy	1725	AAGRGITLDSQVPHLPVLPRTGRTATAMDRLAYLPTADOPSSRHSSPLSPGERTH			1784		
Db	885	--GPKGSLGF-----	PGFGANGKGR--	GTGKRGPRGQSGPTPRGERBPRG	930		
Qy	1785	LTKTTTTSSSEREDRDRERDKREKSILSTTYEHAPIWMPGTBOSSGSSSG--			1842		
Db	931	IT-----		GKPRGKNSGGDGPGRP	950		
Qy	1843	GGGSSSRPASHAHQHSIPRPTDALQORPSVLNHTGKGIITAVEPSKPYLSTSS			1902		
Db	951	GERGPNRGQGTGPRGKPRGPRGKDLPHNROGKGTGQ-----			K 994		
Qy	1903	TSSPVRAATRPRAITHCPGLGTLDGUVYPTLMEVLLPKRADVARPERPRADTHAFLAK			1962		
Db	995	TGPRGPRGVGPRG--	PTGET-----	GPMBERGHPRPG	1026		
Qy	1963	PPAASGLE-----	PASSPKGSER-----	RYLPVPSGHNTIARTPAKN	2001		
Db	1027	PRGEOGLRGVAKKEGTGKDPGRAGLPGKDGPRGLRGPRDGRLPRGVG-----		ALG	1077		
Qy	2002	LAPHNASDPDPAP-----	PASASD-----	PHREKTOSKPSIOLELRSIGYHSGSYS	2049		
Db	1078	LKSGEPRGPRGPRGSPGERPRAGAAGRTGIRGNRPGQRRPGAGE-----		KVPRGK-G	1131		
Qy	2050	PEGYE-----	PVSVPSSPLTHDKGLPKHLEILDKHLSGLRPK-----		2089		
Db	1132	PGGARGRDLQGVPLPRGAPRVGPRGEDQKGEIGBERGQKSKDKKDEQPRPGTBOG			1191		
Qy	2090	---OPGPKLGGEAHNLPHLRPLRPSQ-----		PSSPPLLOTAPGVGN	2129		
Db	1192	PIGQPRGSGADGE-----	PRPRGQGLRGQKDEBSRGRPRGPRGVGLQGLRPRGE		1243		
Qy	2130	QRVVTLAGHIEVITODYTRHNPQLSAPRLAPRLYSPGASCPVLDLRRPSDIYLP--			2166		
Db	1244	KGE-----	IGDVQWMPR--	GRPRGPRGAGADQPR--	QGPFGIGNFGAV	1286	
Qy	2187	-----PRDHAPARA-----		RGSHSSEGRKSRSPRN	2210		
Db	1287	GEKGPREGAERGGLRGEGRGLRGKGERGEKGAQSGAAGPRGPRGDDGPKSPBP-			1345		
Qy	2211	KTSVLGSGEDDGLBPVSP--PEGMTGEGHSRAVYPLLYLDGQETSPRNGSKSPGNTSQRP			2265		
Db	1346	---VGFPGDGPBPBPAGODGPRGX-----		GDDGEGQGVG--	SPGRTGPR	1388	
Qy	2270	AFFSKLTESNAMYKSKOEINKLNTNRNREPEYNIQPGTEIFNNMDALITGTGLMYRS			2322		
Db	1389	-----		GPSGPRGKRGPRG--	PA-----	GPEG	1407
Qy	2330	QAVDEHASTNMGLEAITKKALMKGTDQWDEBSPRLSANAFNPLMASA-----		SLPAM		2381	
Db	1408	ROGEGGAAGEAGL-----		EGPRGKTGPRIGRGAGKRGPRDGLRGITPRPV		1451	
Qy	2382	PITADGRGSDHTLTPGGGKAKVSGRP-----		SSRKAKSPAPGLASGDRPSVSIV		2433	
Db	1452	GEOGLRG-----	SPGRDGPGRGMPRGLPGLKXDSGRGKRGKNGRGLGLIGRGEQG-			1509	
Qy	2434	HSEGD CNRRPTLNRWEDRPSASGSTRPVNPLIMRLQAGYMASPRPPGLPAGSGPLAG				2493	
Db	1504	-EKGP-----		RLPRQSSSG--	PKGEOGITGSPGLGPRGPRGLGPRGPRGA	1545	
Qy	2494	PHNADEPRK				2503	

Db 1550 KGS5GPGPK 1559

RESULT 60

CGH02E

collagen alpha 2(XI) chain precursor - human (fragment)

N:Alternate names: procollagen alpha 2 (XI) chain

N:Contains: proline/arginine-rich protein (PARP)

C:Species: Homo sapiens (man)

C:Date: 07-Jun-1990 #sequence_revision 03-Oct-1995 #ext_change 22-Jun-1999

C:Accession: S34790; A32645

R:Zhukova, N.I.; Brewton, R.G.; Wayne, R.

FEBS Lett. 326, 25-28, 1993

A:Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage

chain.

A:Reference number: S34790; MUID:93314796; PMID:8325374

A:Accession: S34790

A:Molecule type: mRNA

A:Residues: 1-663 <ZHI>

A:Cross-references: EMBL:L18987; NID:G306439; PIDN:AAA5498.1; PID:G306440

R:Kikura, T.; Chesh, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M.;

J. Biol. Chem. 264, 13910-13916, 1989

A:Title: The human alpha2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and gc

A:Reference number: A32645; MUID:89340485; PMID:2760050

A:Accession: A32645

A:Molecule type: DNA; mRNA

A:Residues: 566-1546 <KIM>

A:Cross-references: GB:O04974; NID:G180714; PIDN:AAA52034.1; PID:G180715

A:Note: parts of this sequence were determined by protein sequencing

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (

ed and subsequently O-glycosylated.

C:Genetics:

A:Gene: GDB:COL11A2

A:Cross-references: GDB:119788; OMIM:120290

A:Map position: 6p21.3-6p21.3

A:Introns: 1302/3, 1332/3, 1332/3, 1350/3, 1440/1, 1477/3

A:Note: the list of introns is incomplete

C:Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains (see PIR:CGH

3(XI) chain (see PIR:CGHUC6), initially linked by disulfide bonds among their carboxyl-1-

med with desmosine cross-links made from lysine and allysine residues

C:Function:

A:Description: structural component of extracellular fibrous polymer associated with cell

A:Note: may play a role in controlling the lateral growth of collagen II fibrils

C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology

C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprolin

F:1-154/Domain: non-collagenous (fragment) #status predicted <NC3>

F:1-187/Product: proline/arginine-rich PARP protein (fragment) #status predicted <PARP>

F:255-305/Domain: collagenous, triple helix #status predicted <COL2>

F:306-342/Domain: non-collagenous #status predicted <NC2>

F:343-1356/Region: helical

F:429-431/Region: cell attachment (R-G-D) motif

F:447-449/Region: cell attachment (R-G-D) motif

F:1257-1259/Region: cell attachment (R-G-D) motif

F:1357-1380/Region: carboxyl-terminal nonhelical telopeptide

F:1381-1546/Domain: carboxyl-terminal propeptide (fragment) #status predicted <CTP>

F:1403-1546/Domain: fibrillar collagen carboxyl-terminal homology (fragment) #status acy

F:109-163,1511-1545/Disulfide bonds: #status predicted

F:319/Modified site: alllysine (Lys) #status predicted

F:426,1266/Modified site: 5-hydroxylysine (Lys) #status predicted

F:927,933,1008,1017,1035,1098,1290,1296,1305,1317,1320/Modified site: 4-hydroxyproline

F:1929/Modified site: 4-hydroxyproline (Pro) #status atypical

F:942,1023,1299/Modified site: 5-hydroxylysine (Lys) #status experimental

F:942,1023,1299/Binding site: carboxylate (Lys) (covalent) #status experimental

F:1427,1433,1450,1459/Disulfide bonds: interchain #status predicted

F:1460/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 2.6% Score 344.5; DB 1; Length 1546;

Best Local Similarity 21.2%; Pred. No. 5.4e-05;

Matches 367; Conservative 122; Mismatches 631; Indels 615; Gaps 92;

664 EILQOHKIKJME---KERNARKKKKKAPAAASEAFAFPVVEDEMEASG-VSGNEEMVE 719

Db 1498 PDPVDTPTSSKPSQSKSLKLVDEEEFALRRKRTPSGAKAMHTPKPAVSGEKNIYAFMGT 1557
Qy 1168 KEOGJS-PRGQAGPESLSGVPTPAQENASVLRGTAISVPGSGITTKGIPSRVNSDAITTR 1226
Db 1558 PVQKLDLTENLTGSKRRRLQTPKE-----KAQALBLAG--FKELFQTRGHTE----- 1603
Qy 1227 GSITGTPADVLYKGTITRIIGEDSPRLDRGREDLPKGVHYBEKKHGVSYEGGMSV 1286
Db 1604 -SMTNDKTAKVACKSQPDL--DKNPASSKRLKTSLGK-----VQVKEBLAVG---KL 1652
Qy 1287 TQCSKEDGSSSGPPHE-----TAAPKRTYDMMEGVGRALISASIGLNGRALIPER 1339
Db 1653 TQTSGETTHTEPTGDSKMAFMESPQOILDSASLTG---SKQRLTPKGSKEVPE 1709
Qy 1340 -----HSPHLLKQNHIRGSIIT--OGIPRSYVAQEDYLAREKLLKREGTPPPPP 1389
Db 1710 IAGFIELFQTPSHTKK-----SMTNEKTTKVSYRASQPD-----LVDTPTSSKPP 1755
Qy 1390 SRDLTEAYYQALGLPLKPAHGLVAIVKE--AGRSIHEIPREELRHTPELPLAPPL 1446
Db 1756 KSLKRA-----DTEEEFLAFKQTPSAGKAMH-----TPKRA 1788
Qy 1447 ---KEGSIOTGPTL-KYDTGASTTSGKH-----DVRSLGSPGKTPPP 1486
Db 1789 VGEKQDINTFLGTVPQKLDQPGNLPGSNRRLQTRKEKAQALBELTGFRELFTPTDNP 1848
Qy 1487 V-----HPLDWADA-----PALEPACYEESLSKRGSTASSGSGSIARA 1526
Db 1849 ADEKTTKILCKSPQSDPADPTNTQRPKRSLKKADVEEFLAFKPLPSAGKAMHTPK 1908
Qy 1527 PVIIVPE-----LGKRGQPLTYEDHGAPAGHLPRGSPVWMEPTPLQEGSLSSSKAS 1580
Db 1909 AAVGEKQDINTFVGTPEK-----LILLGNLP-GS--KRRPQTPK-----EKAKYL 1951
Qy 1581 QD---RKLTSTP-----REIA-KSPHSTVPENHPHPIPSPYEHLRGVSGVDL 1623
Db 1952 EDLAGKELFQTPGHTEESMTDOKITEVCSKSPQ--PDPVKTPSSKQRLKISLQKGV 2008
Qy 1624 YASHIPLA-FDPTISIRGILPLDAAAAYLPRHLAPRPYPLHLPPLILGYDPTALENR 1682
Db 2009 KEEVLPVFGKLTQTS-----GKTTQTH-----R 2030
Qy 1683 QTLINDYTSQOMHNTATAMAORADMLRGL-----SPRESSIALNYAAGPRGIIIDLSQV 1737
Db 2031 ETIAGDG--KSIKAFKSAKQMDPRANYGTGMEMKRPTRPEBAOSLEDLAGFK--ELPQT 2085
Qy 1738 P-HLPLVLPPTPGTPATAMDRLAYLPTAQPFSSRHSSPLSPGFTHLTKPTTSSSER 1796
Db 2086 PDHT-----EESTDDTKTKICKSP-----PESMDPTPTST---R 2118
Qy 1797 EEDRODERDREREKSI---LTSTTVHAP-----IMRPGTEGSSGSSGSGGGG 1845
Db 2119 RRPKTPPLGRDIVEEISALKQLTQTHNDKVPGEDEKGINVFERIAKQKLDPAASYTG-- 2176
Qy 1846 GSSSRPASHAHOSPISPRTQDALQORPSVLNNTGMKGI-I-TAVEPSKPTVLNRSTSTS 1904
Db 2177 -----SKQEPRTPKGAQPLEDI--AGLKEILFQTPCTDKPTTHKTT--TK 2218
Qy 1905 SPVRPAATPRATHCPLGLGGLGVPTLMER-----VLLPKAEPVARP- 1948
Db 2219 IACRSPQDPVGT-----PTIFKPGSKRLAKADVEEESIALRKRTPSVGKAM 2266
Qy 1949 EEPRAADTG-----HAFLAKPPASGLEPASPSPKSGEPRLPV-----PVSGHATTA 1995
Db 2267 DTPKPAQGEKQDKAMGTPVQKLDI--PGNLP--GSKRWQTPKEKAQALBLAGFKELP 2323
Qy 1996 RLPAPAK-----LAPHASPPAPAPASADPHREKTQSKPSIOE--LELSLGY 2043
Db 2324 QITPTDTPPTDEKTTIKACKSPQDPDVDPASTK--QRPGRNLRKADVEEFLALAK----- 2377
Qy 2044 HGGSSYPGEVPEVSPVSSBELTHDKGLPKHLEELDKSHLEGELRPQPGPVKLGGEAAHL 2103
Db 2378 ---KRTPSAGKAMD--TPKPAVSDKKNINTFVET-----PVQKLDLILGNL 2417

Qy 2104 PHILRPLPESQESSPLLOTAPGVK-----GHQRVVTLAONISEVITQD-----YTRH 2150
Db 2418 PGGKQFQTPPKAKALBLAGVFKELFQTPGHTESSMTDOKITEVCSKSPQESKTRRS 2477
Qy 2151 HPQOLSAPLPAPLYSPGASCPVLDLRPP--SDLYLPPPD-----HGAAPAGSPHSEGG 2203
Db 2478 SKQRLKIPLVK-----VDMKEEPLAVSKLRTSGEPTQTHTEPTGDSKSIKAF 2525
Qy 2204 KNSPE---PNKTSVUGGED-----GIRP-VSPREGMTEPHSSSAVPLLYRPOE 2250
Db 2526 KESPPQIILDPASVYSGRRQQLTRREKARALBLAGVFKELFQTPGHTESSM-----T 2577
Qy 2251 QTESPMGSKSPGNTSOPAPFSPKLTESNSAWKSGKOEINKLNTHRNBEYVNIQPG 2310
Db 2578 IKNKTIPOCKP-----PPELDTDTSTYCKCKTPKREVEBELAVERLT---QTSQGS 2629
Qy 2311 TEIFNMPAITGTGMLTYRQA-----VOEHAS----- 2337
Db 2630 THTHEPASGDEGIVKQRAKKKNPVVEEBSRRRPPAPREKAQPLEDLAFTLETS 2689
Qy 2338 --TNMGLAIIKALMGKYDQME-ESPPLSANAFPLNASSL-----P 2378
Db 2690 GHTQESLTA-----GRATKIPCESPLEV-----VDTTASTKHLNTRVQKVQKEP 2737
Qy 2379 AAMPITPADGRSDHTLTSPGGGKAKVSGRPSRRKAKSPAPGLASGDRPP 2428
Db 2738 SAVKFTQTSGETTDDADKEPAGDKGIKALKESAKOTPAAPASVYSGRRRP 2787

RESULT 58

S44758

C14B9.6 protein - Caenorhabditis elegans

C1Species: Caenorhabditis elegans

C1Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 09-Sep-1997

C1Accession: S44758

R1FavellO, A.D.

Submitted to the EMBL Data Library, May 1993

A1Description: Sequence of the C. elegans cosmid C14B9.

A1Reference number: S44617

A1Accession: S44758

A1Status: preliminary

A1Molecule type: DNA

A1Residues: 1-1018 <FAV>

A1Cross-references: EMBL:L15188; NID:g289640; PID:g289645

C1Genetics:

A1Introns: 53/2; 74/3; 91/2; 120/2; 336/1; 360/2; 423/3; 535/2; 949/1

Query Match 2.6%; Score 346.5; DB 2; Length 1018;

Best Local Similarity 20.2%; Pred. No. 2.9e-05;

Matches 219; Conservative 143; Mismatches 363; Indels 359; Gaps 43;

Qy 270 ITINQAMRKLLI-----LYFKRRNHAKQKQKFCQYDDQMLELKKVRIENN 319
Db 66 LKVNGLDILLFCSFPCFLFDVLVTFKELARE-----YDKFKAFQDOLKWEET 116
Qy 320 PRRRAKESKVRREYKQPEIRKQRELOERMQSRVQGRSGLSMSAARSEHVESEIIDL 379
Db 117 SERKEYAEFHRVQASSEFELKREKREDRW-----ARERIRGEDKEM 161
Qy 380 SBOENLEKQMR-QLAVIPMLYDADQRIKFINNGLMADPMKYVKDROVNM---WSFO 435
Db 162 LAKENHADKIRLGVAKIPRLITRESRKMDDEFERGSILDKQKHPQGVLDRLSEWSPE 221
Qy 436 EKETREKEMQPKPFGILASLEKTYVAECULYYLTKKNKNYSIVRARSYRRGKSGQ 495
Db 222 ERSLEFSRQADIVKLFHGLTEFPVDKIASDLVLYMNMKKTIDYKDFPKRRVTKYKVG 281
Qy 496 QQQQQQQQQQQQQQPMPSRSGQEK-----DEKEKEKAEK-----BE 533
Db 282 APFSVEBELAFLYFMMPLDSSFPKNSLMCYFCCTRVNGIDLNGTMPKAEYEIFALCPDE 341
Qy 534 EKPEVENDKEDLLKEKTDDTSGEDNDEKAVASGKRTANSQGR----- 578

```

Db      1100 GSPGNIGH-PGSPGLPGKGDKGLPGLDGVGVGVG-----1133
Qy      2154 QLSAPLRLPLVLPFGAGSCPLVLDLRRPEDIYLPR-----PRDHGAPARGSPHSEGGKSPRE 2208
1134 --EAGLPET---PEPTGPAQOKGVPQSD-GIPASAGEKGGVPGKGFPEFPGSK-GDK 1185
Qy      2209 PNKTSV-----LGG-----GEDGIEBVPSEEGTEPGRSRAVYPLLYEDGE---QTE 2253
1186 GSKKEVGRPGLAGSPGIRGVYKGEQGF--MGPRPGQGGGLGTLGTGHNPRVGRKGRGRGQQQ 1243
Qy      2254 PSRMGSKSPGNTSQPPAFFSKLTESNSAMVSKKQOEIN-KLLANTHNREPRYNIISQCTE 2312
1244 PGLPGHPPG---MGPPGF-----PGLNGPKGDKGNQGWPF---GAPGVP 1280
Qy      2313 -----IFNNPATLTGCLMYRQAVQEHASTMNGLEALI-----KRLMG-----KYDOW 2357
1281 GPKCDPGFQGGPFIIGGSPGII-----GSKGMGLPGVGFQGGKGLPGLQGVKADQG 1332
Qy      2358 EESPPLSANANPLMASASLPAAMPITAADRSHTLLSPCGGGKAKVSGPSSRAKASP 2417
1333 DQG-----VFQPKLIQGPFGPGPYDIYIKG 1357
Qy      2418 APGLASGDRPSPVSSVHSEGDONRRTPLTNEV-----WEDRPSSAGST--PF--PYN 2465
1358 EPGLPBPGCPGGLKGLQGRPGPKQGGVTSVGLPGRPGVPGFGARPGQKQKETSPPFGPBG 1417
Qy      2466 PLIWRLOAGVNASPPPPGCLPBGSGFLAGP-----HHAMDEEKPRLIC 2507
1418 P-----RGFGPGRPGDGLPGSMGPPGTPSVYDHGFLVTRHQQITDPD---LC 1460

```

[illegible]

RESULT 57

2

cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
C:Accession: A48666
R:Schuster, C.; Duchow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdes
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-
ous.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SOCH>
A:Cross-references: EMBL:X65550, NID:g415818, PIDD:CAA6519.1, PID:g415819
C:Superfamily: Kinase-Interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/DNA: kinase interaction domain homology <KIH>

```

QY      631 SE-EAAPPVEDEMEAS-GVSGNEEENVEEAELHAG-----NEVPRGECSSPATYN 743
Db      1010 LQPEINTPTHTKQOLKASLGVGKEELAVGKFTRTSGETHTHREBAGDOKSIRTEK 106
QY      744 NS-----SDTESI-----PSPTTEAKD--TGONGP 767
Db      1070 ESPKQILDPAARYTGMKKMKRPTPKREAGSLBIDLACIKELFQTPGPEESMTBCKTIKINC 112
QY      768 KKPATLGADGPPGPPTPPRTSRAPTEB-----TPASATGATPPPPAPSPSAPP 819
Db      1130 KSPPEASVDTPSTQWPKRSIRKADVEEELFLATRLKLTPE-SAGKAMLTPEKPA----- 118
QY      820 PVVPEEKEEETAAAPVE-----EGEBQKPPAAEBELA-----VDTGKA 858
Db      1181 ---GGDEKDIKAEWGTVPQKLDLAGTLPGSKQLOTPPKKCAQLBIDLACIKELFQTPGHT 1233

```

Query Match 2.6%; Score 348; DB 2; Length 3256;

2

Matches 537; Conservative 352; Mismatches 1129; Indels 852; Gaps 133;

2

4 STOLVAQTRATEPRYPHSLSYVQIARTHTDVGLELEYQHHSRD-YAS----- 51

✱

Db 325 SVQTPSKAVGASFPLYEPAKMKTPIVQYSQQQNS---PQKHKNKDLVTTGRRRESVNLGKS 380

2

QY 52 -----HLSPGSIQPRRPSL-----SEFQPGNERSQELHLPESHSLP---EL 95

*

Db 381 EGFKAGDKLTTPRK-LSTRNRTPAKVEDAADSATKPENLSSKTRGSIPTDVEVLPTETEI 439

2

QY 96 GKSEMEFIESKRPLELPDPLRPSPL-----LATGPAGS-----EDLT KD RSL 141

2

Db 440 HNEPFLTLWLQVERKIQKDSLSKEKLTGTTAGQMCGLPGLSSVDINNFGDSINESEGI 499

QY 142 TGLEPVSPSPHTDEL--ELVPR-----LSKELIQNMDRVREITMVEQOI 190

3

db 500 PLKRRRV - - FGGHLRPELFDENLPPNTPLKRGAPTKRKS LVMHTPPVLKKIKEQPQ 557

QY 191 SKLKKQQLLEE-----AAKPEPEKVPSPPIESKHSVLQIYDENRKAEEAHRI 244

QY 1479 SPGRTPPVHLDVMDADRALEACYSERLSKRP-GTASSSGCSARGAPVIVPELGR 1537
 Db 1353 GMGQWLPAASGADPLCRNPA-SRSLKGLNLSKSLAADCPKAEAFNSPTLSLGLAR 1411
 QY 1538 QSPFLYEDHG-----APPAGHLPR-----GSPVTRKEPPRLQ--EGSLSSKASQDRKLT 1586
 Db 1412 -APKRSKSKGTGLGPEKPLEKPPCGQLL-APHDRASSVQGGEDNDSGGGKK- 1467
 QY 1587 STPREIAKSPHSTVPEHHPHDIPSYEHLRLGVSGVDLYRSHIPLAFDPTSLPRGILPDA 1646
 Db 1468 --PTEBELGPASQPEGRP-----CQOTBAQKQPCQAS 1499
 QY 1647 AAYLPLRLAENPPTPHLYPPLVLRGYDPTALENRQTIINDYITSQOMHNTATAMA-- 1704
 Db 1500 YSSYSKR-----KRLSGRGKTA-----HASPCKGRYR 1528
 QY 1705 QRAMLRLGLSPRESSIALNVAAGPRGIDLSQVPLPVLVPTPTGTPATMDRLAYLPTA 1764
 Db 1529 RROQVPLPLDAPEIRLKYISSCKRLADSRTP-----A 1563
 QY 1765 POPFSRHSPLSPGPFTHLTKPTTSSSRERDRDRDREREKSLTSTTV-- 1821
 Db 1564 FSPFV-----RVEKDAYTITCTVNS 1585
 QY 1822 --EHAPTRPCTEOTSSSSSGSSG--CGGSSSRPASHAHQHSISPRTQALQORP 1875
 Db 1586 PGDEPRKPMKSSSAASSSTSSSLPAGASLITFPFG-----SVLQGRP 1630
 QY 1876 SV-LHNTGMKGLITAVEPSKPTVLASTSSPVPRPATFPATPCLGLTLDGVPTLME 1934
 Db 1631 SLPLSTHMLG-----PVSGALSTSLCY-CCLCONRANFDLGLCPYTP---- 1676
 QY 1935 PVLLPKAPRYARPERPRADTGHAFIAPKAPASGLEPASPSSKSGSEPPRLVPSGHATI 1994
 Db 1677 EHCLPKKKPKLKEKARLEGTLEASLPLERTLKGLCASATTA--PTATITTPPAL 1733
 QY 1995 ARTPAKLAIPHASDPAPAPASADPHREKTKQSPFIOELSLSLGHSYSPEVE 2054
 Db 1734 GRL-----SRDGPADPAK-OGPLRTSARGLSRRLQ-----SCVCCDG- 1770
 QY 2055 PVSPVSPSLTHDKLPKHLLELDKS--HLBGLRPPKPGPVKLGGAHLPH 2105
 Db 1771 -----OGDGG--EIVAQADSKRHKCKEKAFTPEG-----GDTEHWH 1807

RESULT 56
 CGMS4B
 collagen alpha 1(IV) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000
 C:Accession: A33525; S01454; A28066; A02864; A25636; A23301; S19079; A32003; A31766; S19
 J:Mutlukumar, G.; Blumberg, B.; Kurkinen, M.
 J: Biol. Chem. 264, 6310-6317, 1989
 A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Diff
 A:Reference number: A33525; MUID:89197932; PMID:2703490
 A:Accession: A33525
 A:Molecule type: mRNA
 A:Residues: 1-1669 <MUT>
 A:Cross-references: EMBL:J04694; NID:9556296; PIDN:AAA50292.1; PID:9556297
 R:Wood, L.; Thieriault, N.; Vogel, G.
 FEBS Lett. 227, 5-8, 1988
 A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the al
 A:Reference number: S01454; MUID:8811221; PMID:3338568
 A:Accession: S01454
 A:Molecule type: mRNA
 A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'P', 404-480, 'L', 482-492, 'H', 494-71
 A:Cross-references: EMBL:X06777
 R:Killem, P.D.; Burdello, P.; Sakurai, Y.; Yamada, Y.
 J: Biol. Chem. 263, 8706-8709, 1988
 A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen cha
 A:Reference number: A28066; MUID:88243724; PMID:3379041
 A:Accession: A28066

A:Molecule type: mRNA
 A:Residues: 1-129 <XII>
 A:Cross-references: EMBL:J03758; NID:9192669; PIDN:AAA37439.1; PID:9192670
 R:Oberbauer, I.; Laurent, M.; Schwartz, U.; Sakurai, Y.; Yamada, Y.; Vogel, G.; Voss,
 Eur. J. Biochem. 147, 217-224, 1985
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A:Reference number: A02864; MUID:85127033; PMID:2578961
 A:Accession: A02864
 A:Molecule type: mRNA
 A:Residues: 1276-1669 <OBE>
 A:Cross-references: EMBL:X02201; NID:950233; PIDN:CAA26132.1; PID:91333876
 R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogel, G.
 Gene 43, 301-304, 1986
 A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
 A:Reference number: A25636; MUID:86301886; PMID:33755692
 A:Accession: A25636
 A:Molecule type: mRNA
 A:Residues: 1149-1396, 'S', 1398-1424 <NAT>
 A:Cross-references: EMBL:M14042; NID:9192286; PIDN:AAA37342.1; PID:9192287
 A>Note: the authors translated the codon CAG for residue 1374 as Arg
 R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Finlay
 J: Biol. Chem. 262, 8496-8499, 1987
 A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
 A:Reference number: A94680; MUID:87250460; PMID:3597383
 A:Accession: A29301
 A:Molecule type: mRNA
 A:Residues: 1441-1669 <KUR>
 A:Cross-references: EMBL:M15832; NID:9192282; PIDN:AAA37340.1; PID:9387115
 R:Killem, P.D.; Burdello, P.D.; Martin, G.R.; Yamada, Y.
 J: Biol. Chem. 263, 12310-12314, 1988
 A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
 A:Reference number: S19079; MUID:88315019; PMID:2842328
 A:Accession: S19079
 A:Molecule type: DNA
 A:Residues: 1-28 <K12>
 A:Cross-references: EMBL:J03944; NID:9192673; PIDN:AAA37442.1; PID:9466503
 R:Kayes, P.; Wood, L.; Thieriault, N.; Kurkinen, M.; Vogel, G.
 J: Biol. Chem. 263, 19274-19277, 1988
 A:Title: Head-to-head arrangement of murine type IV collagen genes.
 A:Reference number: A92702; MUID:89066738; PMID:3198626
 A:Accession: A32003
 A:Molecule type: DNA
 A:Residues: 1-28 <KAY>
 A:Cross-references: EMBL:J04448; NID:9192666; PIDN:AAA37437.1; PID:9450449
 R:Burdello, P.D.; Martin, G.R.; Yamada, Y.
 Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promo
 A:Reference number: A94220; MUID:89071759; PMID:3200851
 A:Accession: A31766
 A:Molecule type: DNA
 A:Residues: 1-28 <BUR>
 A:Cross-references: EMBL:M23333; NID:9340878; PIDN:AAA51625.1; PID:9535668
 R:Sakurai, Y.; Sullivan, M.; Yamada, Y.
 J: Biol. Chem. 261, 6654-6657, 1986
 A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
 A:Reference number: S19094; MUID:86196099; PMID:3009468
 A:Accession: S19094
 A:Molecule type: DNA
 A:Residues: 1110-1135, 1189-1316, 1342-1383, 1418-1487 <SAK>
 A:Cross-references: EMBL:M13027
 R:Schuppan, D.; Timpl, R.; Glanville, R.W.
 FEBS Lett. 115, 297-300, 1980
 A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
 A:Reference number: S16909; MUID:80246483; PMID:6772473
 A:Accession: S16909
 A:Molecule type: Protein
 A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-123
 R:Schuppan, D.; Glanville, R.W.; Timpl, R.
 Eur. J. Biochem. 123, 505-512, 1982
 A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amin
 A:Reference number: A25991; MUID:82186723; PMID:6804236
 A:Accession: A25991
 A:Molecule type: protein

Db 937 -FTKSENSNRTCSMPYAK---LEYKSSN-----DSL-----NSVSSSDGY 973
Qy 320 PRRAKSKVREYER-----QFP-EIRKORELOERQSRVQSGSLMSARSE 369
Db 974 GKRQMKPSISYSYEDDESKFCYSQGYPADLAHKHSANHMDNDGELDTPTINSLKXSD 1033
Qy 370 HEVSEIIDLSEOELEKOMQOLAVIPMLYDADQORIKFINNGMLAMPKMYKROVM 429
Db 1034 -----EOLNSGRQ-----SPSQNERMARPKHIIIDELIKOSEQORSR 1069
Qy 430 NMMSEOE--KETFEKFMQHPRNGLIASFLERKTAVACVLYUYLTKNENYKSLVRSY 487
Db 1070 NQSTTYPYTESTDDKHLKQPHFG-----QOCVSPYSRGANGSETNRVGSNH 1119
Qy 488 RRRGKSQQQQQQQQQQQQQQQPMRPSROEKDEKEKE-----AKKEEKEPVEND 541
Db 1120 ---GINOVVSQSLQOEDYEDDKPTNYSERYSEEOHEEERPTNYSJKYNEBKRWQOP 1176
Qy 542 KEDLKEKTD-----DTSGEDN-DEKAVASGRKRTANSGRRKRITRMA 587
Db 1177 IDYSLKATDLPSSOKOSFSPSKSSSQSSKTEHMSSESTSTPSSNAKQNLHPSA 1236
Qy 588 NEANSEEAITPOQSALASMEINESSRWTEBEMETAKGLLEHGNMGAIRAMVSKTVS 647
Db 1237 QSRSGQ---PQKATCKVSSINOETIOTYCEDTP-----ICFGRCSLSLS 1281
Qy 648 QCKNPFYNYKRRQND--EIIQOHLKMEK--ERNARKKKKAPAAASEEAPPPVEDE 703
Db 1282 SAEBEIGCQTOEADSNATLQIAEIK-EKIGTRAEADPVSEVPVSOH-----PRTKSS 1335
Qy 704 EMEASGVSGNE---EMVEEABALHASGNEVPRGE-----CSGPATVNNSSDSEIIS 753
Db 1336 RLQSSLSSEBARHKAIVFSSGASPSKSGAQTSPSPENHYOETPLFMSCTISYSDS 1395
Qy 754 PHTEAKTQNGKPRPATLGADP---PPGPPTPRRTSRAPIEPTPASEATGATP 808
Db 1396 FESSIASVSEPCSGMVSGIISPSDLPSGCGMPSPRSK-----TP 1439
Qy 809 PPAPSPAPPPVPVPEKEKEBETAAPVEEGEOKPPRAEBLAVDTGKA---EEPVKSE 865
Db 1440 PP-PPQOTQOTREVPKMK-----APPAEKRESGPKQAAVAAVQOVLPDADTLHF 1491
Qy 866 CTEBAEBC-----PAKGDAEABEATAEGALKAKEKGGSGRATTA---KSGAPQDS 915
Db 1492 ATBSTPDDSCSSSISALSLDBEPTQKDELKIMPVOENDNGNTESEOPKESNENOK 1551
Qy 916 DSSATCSA--DEVDEABEGD---KNRLSPRPSILPTGDPRAVASPOKPLDKQLKOR 969
Db 1552 EAEKTISEKULDDSDDDIEILEECIISAMPT-----KSSRKAKKP---AQT 1597
Qy 970 AAALPPIQVTVKNEPRREDAAFTKAPAPAPPPQULQPSDAPQOPSS--PRGKSRSAP 1028
Db 1598 ASKLPAPAPKPSQULPVYKLLPSQ-----NRLQOKHVSFTPGDMPR----- 1640
Qy 1029 PADKAPAAABQKULPGDPFCMTSGLPFPVPPREVIKASPHAPDPSAFVAPRGHPLPL 1088
Db 1641 ---VYCVBEG-----TPINFTTA-----TSL 1657
Qy 1089 HDTAAPVLPAPPTISNPPLISSAGHPSELERQIGALSQMSVQLHVVYSEHAKAPVGP 1148
Db 1658 SDL-----TIESPP-----NELAAGBVGWGAQSGEKKDITLPT 1692
Qy 1149 TMGLPLPMDPKLAPFSGVQKQOLSPROQAPPELSGP-----TAQASVLRGTALASV 1203
Db 1693 -----BGRSTDEAQGGKTSVITIPELDNDKAEBGDILAEICINSAM 1732
Qy 1204 PGG-----SITKGIPSTRVPSDSAITR-----GSITGTADVLVYKGTIRI 1246
Db 1733 PKGSHKFRPVYKTIINDVOQOASASSAPYKNOLODKKKKPTSPVPIQONTETRRVVK- 1791
Qy 1247 IGEDSPSLDBGREDSLPKGHVITYEGKKGHVLSEYSGMSVTQCSKD---GRSSGPPHE 1303
Db 1792 -NABSKNNLMBRVS DNK-----DSKQONLKNNSKDFNDKLPNNEDRVRGSGFADSPHH 1845

Qy 1304 TAAPKTYDMMEGVRGAISASIEGL-----MGRATPRERHSPHHLKEQHINGSI 1355
Db 1846 ---YTPLEGPPYCFPSRNDLSLSDPDDDDVLSR---EABELKAKENKESEAKV 1894
Qy 1356 TOGIPREYAEQEDYLREAKLIRKEGTPP-----PPSRDL-----TEAVYKQ 1400
Db 1895 TSHTELINSQOSNAKQALAKQPINRQOPKILQKQSTFPQSSDIDPRGAATEKQONF 1954
Qy 1401 ALGPLKLPRAHEGLVATVKEAGRSIHEIPREBLNHT-----PELPLAPRLKEGSI 1452
Db 1955 AIENTPVCFSHNSLSLSLSDIDQENNNKENPEIKETERPDSQGPSPKQASGVAPKSFV 2014
Qy 1453 OGTEPKYDTGASTG-----SKKHDRVSLIGSPGR 1482
Db 2015 EDTPVCFSRNSLSLSLSDISEDDLQOCISAMPKPKPSRLKQDNKHSFRNNGJLGE 2074
Qy 1483 TFPVPHPLDVADARALERACYEESLKSRRPCTASSGGSIRARGAPVIVPELGRPROSPLT 1542
Db 2075 ---DLTLDKDIQRPQSEHGLS--PQSEMFQWKALQEGANSIYSSL----- 2115
Qy 1543 YEDHAPRAGHLP-----GSPVTMRBPTPRLQOEGSLSSQASODRK 1584
Db 2116 ---HQAAAACLSQASDSDSLILSKGISLSPFHL---TPQOEKPFYSNKG--- 2164
Qy 1585 LSTPRELAKSPHSTVBNHHPHISPYEHLRGV-SGVDTLRSHIPLAFDPTSIIPGLPL 1643
Db 2165 ---PRILKPEKSTL-----ETKLTSESKGIGKGGKVVKS----- 2197
Qy 1644 DAAAYVLPRLHLPNPTYPHLPYLLRGVYDPTALENRQTLINDYINSQQMHNHTATAM 1703
Db 2198 -----LITG-----KYRSNSELISQW----- 2213
Qy 1704 AQRADMLRGSPRESSALNLYAAGRGITDLSQVPHLPVLPPTPGPATMDRLAVLPT 1763
Db 2214 ---KQPLQANMPSISRG---RTMIIHIGV----- 2236
Qy 1764 APQFSSSHSSPLSPGCPHTLTKPTTSSSERDRDRDRERERKSLTSTTT--- 1820
Db 2237 ---RNSSSTSPVSKKGP--LKTTPASKSPBEGQ-----TATSPRG 2273
Qy 1821 ---VEHAPMPGTEQSSGSGSGSGGSSSPRASHAHQNSPISPTODALQOR 1874
Db 2274 AKPSVSELSVPAQ--TSQIGSSKAPBSGSRSTSPRAQOLSRITQSPGANSI--- 2329
Qy 1875 PSVILNTEMKGITAVESKPTVLASTSTSPVRPATFTPATHCPLGCTIDGYR--- 1930
Db 2330 ---SPGRNGI---SPPNKLSQLPRTSSPSTA-----STKSSGSGKMSYTSFGRQW 2373
Qy 1931 ---TMEPVLLPKKAPRVARBERPRADTGHAFLLAKPPARSGLEP-----ASSPSKSE 1980
Db 2374 SQONLTKQGTGSKANSST---PRSEBASGKLQOMNNGNANKVLELBSMSSTSGSE 2428
Qy 1981 ---PRPLVPVSGHATITARTPA--KNLAPHNASDPAPRAPASDPAREKTQSKPFSIOE 2035
Db 2429 SDRSERVUVQSTFIKEAPRTIARKLEBASLESLSPPSRAPFRTSQOQT--PVLSPS 2487
Qy 2036 LELNSLGYHSSYSPEGEVPSVSSPSGLTHDKGLPKHLELDKSHLEGLRPRQOPGVK 2095
Db 2488 LPDMSLSTH--SSVQAGGRKLPRLNISPTIENDGRAPARNDIARSHSHPSSL---PIN 2542
Qy 2096 LGG--EAAHLPHLAPLPSPQSSSPLQTAGVGNHQRVUPLAHQISIVLITQDTRHHPQ 2153
Db 2543 RSGTWKREHSHX-----SSSLPRVSTWRTTSSSSLSASSESSKAKASEDEKH--- 2591
Qy 2154 QLSAPLPALYSPGASCPLVLDLRRPSDLYLPDPDHGAPARGSPHSEGRSPENKTS 2213
Db 2592 -----VASIGTK-----QSKENGVSAKGWTRKIKENEFSTNBS 2627
Qy 2214 ---VLGGEDGIE-----PVSPPGQMTPEBGRSAVLPVLYLRDGBQTEPSMGSKSPONT 2265
Db 2628 QTVSSGATNGAESKTYLLYQMAPAVSKTEDVWVRIEDCPI-----NNPRSGRSPTGNT 2679


```

QY 1809 EREKSLTST-----TVEHAD-----IWRPGE-----QSS-----GSSGSGGCGG 1846
Db 2162 TSSLASVSTSAEPTGASVAVKLSLSPDVSQSTKTFDATTESSTVOASSTSGTSEVK 2221
QY 1847 SSSRASH-----SHAHQSP-SPR-----TODLAQGPVYLHNTGMKGIIITA 1889
Db 2222 STSEESHVTKLITSSNPSVSVPTSPKSTPTVEBSTEOPSTTSPGSLTPMNSSEV 2281
QY 1890 VEPSPKTVLSTSTSPVPAATFP-----PATRCPLG-----GTLGVPYPTL 1932
Db 2282 LITSEPHVL-SSLSLSDVVSQSTTPPNLSESTVEPEKTSVSVLNSEBPTTEA--FTT 2338
QY 1933 MEPLVP-----KEAPVAPRPRADTGHAFLAKPPARSGLEPAS-----SFS 1976
Db 2339 LSPDLITSTNNLSQSTVSTEDREISENS--EKPTAPRLVTSVTHVASSSPDVT 2396
QY 1977 KSEPRPL-----VPVSGHATTARTPAKNLAPHNASPPDPAPASADHREKTOGK 2029
Db 2397 ESSEPDLLGSSSTENIPEASKOTISSTP-----TPD-----TTTSEPTSTMS 2443
QY 2030 PPSIOELRLSLGCHGSSVSPGCEPVSPSP-----SLTHDKGLPKHLEELDKHLEGE 2085
Db 2444 P-----DLSTTSNVLSSSSTTPES--SEKSPVSSSTEGISVSTEFKVESTISSVLEBD 2498
QY 2086 LRPKQGPVYKLGSEAAHLPHLRPLPSQSSSPLQTA--PGVKGQRVTLAQHT---S 2140
Db 2499 LTKTTPSP-LEETTTASETSEPLTEDSLTVSVRIHE-LTTSSENVPKES 2546
QY 2141 EVITQDYTHHPQOLASPLPAPLYSPGASCPVLDLRPPSDLYLPPDHG--APARGSPH 2199
Db 2547 ESTTSSSESKPQEPAGILITSVVPTSSVLSITSEIALITSNTPFGQGRPTITSPK 2606
QY 2200 -----SREGKRSPEPNKTSVLGGGEDGIEPVSPGKTEPEHGSANVPLYLRDGEOT 2252
Db 2607 SLVKTSTSPVTVSSPSESTKRTVSTVSTTPEETTT---SESLI--LTAAPSKPT 2661
QY 2253 EPRMSKSPGNTSQPAPFSKLTESNSAMVMSKKOEINKKLTNHRNEPVNI-----S 2307
Db 2662 ESTTESSEAP---TTPAKTSETPKPNVSTSRKSTE-NVETISQSGSLESSTMSSTS 2716
QY 2308 QPCTEIFNMPATITGTMTRSOAVOEHAFT-----NMGLEAIRKALMKYDQWEE 2359
Db 2717 EPET---NAPAVTVSSEAS--STLLENSTSTPTSEASVLSLFPESITSEAVTVGS 2771
QY 2360 SPP-----LSANAFNPLNASASLPAPMTADRSHTLTPGCGGKAKVSPSR--- 2412
Db 2772 RAFAETMSSESHREISTVSESEPEIPLSTTVSPVTVTA-----SSIPSEPTL 2822
QY 2413 ---KAKSPAPGLASGDRPVSVSSEBDCNRTPLTNKVMEDRPSASGTFP--PYNPL 2467
Db 2823 SSVTSSSTPRVRLITGTPDLIVSVTVPSHGRRKONIT-----ASVVSNGTSPT 2872
QY 2468 IMRLQAGVMASPPP---PGLPAGSGELAP 2494
Db 2873 ILPSESLTTPQPPPTTTAKPATTSKGRP 2903

```

RESULT 54

RBHUP

adenomatous polyposis coli protein - human

N/Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 21-Jul-2000

C/Accession: A37261, B39658, A4928, A49319, I54271

R/Kinzler, K.W.; Nishisho, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith

chui, S.F.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.

Science 253, 661-665, 1991

A/Title: Identification of APC locus genes from chromosome 5q21.

A/Reference number: A37261, MUID:91335210, PMID:1651562

A/Accession: A37261

A/Molecule type: mRNA

A/Residues: 1-2843 <KIN>

A/Cross-references: GB:M74088; NID:g182396; PIDN:AAA03586.1; PID:g182397

```

R/Joslyn, G.; Carlson, M.; Thliveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Grod
arrington, J.; McChesom, J.; Wasmuth, J.; Le Paslier, D.; Adderham, H.; Cohen, D.; Le
Cell 66, 601-613, 1991
A/Title: Identification of deletion mutations and three new genes at the familial polypc
A/Reference number: A39658; MUID:91330307; PMID:1678319
A/Accession: B39658
A/Molecule type: DNA
A/Residues: 1-183, 'L', 185-969, 'N', 971-1308, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P'
A/Cross-references: GB:M73548; NID:g190163; PIDN:AAA6054.1; PID:g190164
R/Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelst
Cancer Res. 52, 643-645, 1992
A/Title: Disruption of the APC gene by a retrotransposon insertion of LI sequence in a c
A/Reference number: A44928; MUID:92119623; PMID:1310068
A/Accession: A44928
A/Molecule type: DNA
A/Residues: 1506-1525 <MTK>
A/Cross-references: GB:S78214; NID:g243541; PIDN:AB21145.1; PID:g243542
A/Note: sequence extracted from NCBI backbone (NCBI:78214, NCBI:78218)
R/Splitz, L.; Olschansky, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelber
Cell 75, 951-957, 1993
A/Title: Allele of the APC gene: an attenuated form of familial polyposis.
A/Reference number: A49319; MUID:94073973; PMID:8252630
A/Accession: A49319
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 'G', 143-171, 'P', 173-179 <SPI>
A/Cross-references: GB:S67787; NID:g461061; PIDN:AAD13997.1; PID:g461697
R/Lambertz, S.; Ballhausen, W.G.
Hum. Genet. 90, 650-652, 1993
A/Title: Identification of an alternative 5' untranslated region of the adenomatous poly
A/Reference number: I54271; MUID:93186137; PMID:8383094
A/Accession: I54271
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-4 <LMB>
A/Cross-references: GB:S56365; NID:g266243; PIDN:AAD14918.1; PID:g4262770
C/Genetics:
A/Genes: GDB:APC
A/Cross-references: GDB:119682; OMIM:175100
A/Map position: 5q21-5q22
A/Note: mutations of this gene can result in familial adenomatous polyposis or sporadic
C/Superfamily: adenomatous polyposis coli protein
C/Keywords: cancer; familial adenomatous polyposis; tumor suppressor
F.1-730/Domain: leucine-rich <NTD>
F.7-72/Region: coil #stratus predicted
F.185-227/Region: coil #stratus predicted
F.731-2832/Domain: serine-rich <CTD>
F.1131-1156/Region: acidic
F.1558-1577/Region: acidic
F.1866-1893/Region: highly charged

```

Query Match 2.7%; Score 350.5; DB 1; Length 2843;

Best Local Similarity 17.8%; Pred. No. 6.5e-05;

Matches 473; Conservative 357; Mismatches 1007; Indels 82; Gaps 109;

```

QY 50 ASHLSPGSLIOP--QRRPSLSEFO-----PGRSGQELHLPESHY 91
Db 740 ANIMSPGSLPSLHVRKQKALFAELDAQHLSFTPNIDNLSKASHRSKQRLKQSLYGDY 799
QY 92 LELGKSEMEFIESKRPRLLELPRLRSPPLAQCPAGSFDL---KDRSL----- 141
Db 800 VEDTNHNDNRSDNFTGMVTLSPLNTTVLPSSSSSSGSLDSRSRSEKDRSLERBGIG 859
QY 142 TGLKLEVPSPSPPHPTDPELEVPRLSKELLQNDMDRVDRITVWEQJSLKKQOOLE 201
Db 860 LGNYHPTATN-----PGSSSKGL-----QISTAAQIAKV-----ME 892
QY 202 EBAKPEPEKVPSPPIESKRSIVQIYDENR--KAAEAARLIEGIGPVEVLEPLVQ 259
Db 893 EVSAIHITQEDRSSGSTB-----LHCVTDERNALFRSSAAH-----THSNVY- 936
QY 260 PSTROYENIKINQMRKLLIYFKRRHAKQKQKCCQYVDQLEALBKVERIENN 319

```

D _b	1813	G G G S V F G G T S A A T T T A A T T G C F P C Q A G F G S S N T G S V F G A A S ----- T G G I V F G Q S	1866
O _y	2428	P S V S - S Y H S E G D C N R	2441
D _b	1867	S S S G S V F G S G A N T G R	1881

RESULT 53

hypothetical protein ZK783.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513

submitted to the EMBL Data Library, August 1994
 A:Description: The sequence of C. elegans cosmid ZK783.
 A:Reference number: 221536
 A:Accession: J134513
 A:Status: Preliminary; translated from GE/EMBL/DD83
 A:Molecule type: DNA
 A:Residues: 1-3507 <FAV>
 A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
 A:Experimental source: strain Bristol N2; clone ZK783

A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match	2.7%;	Score 354;	DB 2;	length 3507;
Best Local Similarity	19.1%;	Pred. No. 6.1e-05;		
Matches 4/6;	Conservative 288;	Mismatches 879;	Indels 848;	Gaps 117;

```

Qy      550  TDDPISGENDDEKEAVASKGRKTANSQGRKRGIRTRSMANSEANSEAIITQOASAEIASMWL 609
          :      :      :      :      :      :      :      :      :      :
Db      715  TEDEGEDEDLMEGSGSGMSSTINGT-----ITGSPREEGTIRVIT-----TL 759

```

```

QY      610 NESRWTEEMETA-KKGLLEHGHWMSAIARWWSKTVQCKNFYFNKKRQNDLEILO 668
          | | | | |
Db      760 GE-----DOEPETATKPGISAPDKTGE-----GSKTSD-----GE 790

```

Db

	669	HKLMEKE-----RNARRKKKKAAPAAASEEAAFPVVEEDBEMEKESGVSNSEEBENV	718
OY		:	
	791	EKLTVKCKGKAQSGSSSATSSGKKSEATISGSSSSA-----KSCTGSEASGSSGC-----	840

QY 989 --AAPTKAPAPPPQUNQPE--SDAPQPGSSRGKGRS.PAPADKFAAQAQKLP--G 1044

1232 TSGVDGKPTTPAPPPSSSAEASTRIPTTSEAPBEGGGEAGVESPQSGGSESTAPDG 1291

```

QY      1045  DPPECTSGLPFPPVPREVIKASPHAPDPSAFSYAPRGHPLPLGHDTPARVLPAPPT--I  1102
Db      1292  VSP--TSSATAPEVPTTSASSTPAVBSG-----IPSTSKPTAPLETTAPSTEV  1340

```

QY	1103	SNP-----	PLISAKHPVLERIGALISQMSVQ--	LHVPSEHAKAPVGPPT	1143
Db	1341	ITPFGSGTEESTLPTEGSGE----	STTSAPATVPATVLPQNNKEKPTKOT		1391

Qy 1150 MGELPMDERKTLAFSGVKEQQLSPRGAG-----PEPSLGV-- 1186

Db 1392 FALPTTTGAPQANDSSVENTKCTSSDDEGDLALCERRTGVCRCEPFGAPPPKKSVDV 1451

Qy	1187	-----PRA-----	-----QASVLRGTALG-----	1201
Db	1452	DECATGDHNCHEASARQNVGVGACFCPTGFRKADDSQDIDECTEHNSTCCGANAKCV		1511

Db

```
1512 NKPGTSCCECENGLGDGYQCVPTTKKPCDST---QSSKSHCSBSNMSCVDTVDGSGVBC 1566
```

Db

```
1569 KECMGYKKSGKVCEDINECVAEKAPCSLNANCVNM--NGTFSCSCKGG---YRGDGEF 16422GF
```

Db 1623 CTDINCEDEHRHPCPHAECTNLLEGSFKCECHSGEGDGIKKCTNPLERSCEDVEKFCGGRV 1682

Db 1683 DHVSLSVRIYNGSLSSVCECEPGRFRFEKESNSCVDIDECESRNNCDDPASAVCVNTEG-17411

Db
1742 --STRCEAGY-----:||:|:|
-----EGBG-----:|:|:|
-----GCTDIDECDRGAGCDSMAMCINRMGSCGC 1786

Db 1787 KCMAGYTSGATCIKIEEFPKSDKTACIDEMSRCLCELEKQCTVDEEVEVPQ----- 1837

Db 1838 -----CGACLPQHHPINGTCQSLQISGLCAQKNDCKNHAECIDIHPS 1880

Db 1881 HFCSGCPDGFIGDGMICDDVDECNAGMODDENTKCENTIGSFNVCVLGGPKYDEKCVVD 1940

[illegible][illegible]

```

Db      2019  -----PEVSTSSSKXTTAAETTIVSTSPSESSSSSEAPLTSSP 2054
Ov      1701  --TAAORADMTREGISPRE--SGLATVNAAGPGITDI/SOVH/LPVI/VPTPTGPTATM 1755

```

```

Db      2055 ATTTEVITESSVXSTTPKRESSSEITVXKLSKQPEVTESS-----VKSSBPTPTS 21066
Ov      1756 DRLAYLPTAPOPFBSR--HSSSPSPGPGPTHL-----TKPTTSSSERERDRDRERDRDR 18088

```

Db 2107 QSVI--STVPEKSTVLSSEAPVTSTSTPEVHTSSEIKPSLSASS---TTGDTNSTTPS 2161

Db 2199 K-----NE-----MFRKEILRR-----LDLITLLEL-G 2220
Qy 2353 KYDQWEEPPPLSANAFLNPLNASLIPAMPTIADGRSDHTLTSGGGGA--KYSGRPS 2410
Db 2221 AAEDEDOKRDLK-----QIP-----TSEEDTDSKADSMGAEASAFRRILSRSS 2264
Qy 2411 SRKAKSPAGLSDGRPPSVSSVSEG 2437
Db 2265 TWGNNGSPS-ASGTTSPSTSSISGG 2290

RESULT 52
S26058
Probable transforming protein (can) - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S26058
R: von Lindern, M.; Fornerod, M.; van Baal, S.; Jaegle, M.; de Wit, T.; Buijse, A.; Grosveld, M.; Cell. Biol. 12, 1687-1697, 1992
A:Title: The translocation (6;9), associated with a specific subtype of acute myeloid leukaemia.
A:Reference number: S26058; UID:92195315; PMID:1549122
A:Accession: S26058
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2090 <VON>
A:Cross-references: EMBL:X64228; NID:g29652; PIDD:CAA45535.1; PID:g29653

Query Match 2.7%; Score 355; DB 2; Length 2090;
Best Local Similarity 20.6%; Pred. No. 3.2e-05;
Matches 362; Conservative 177; Mismatches 656; Indels 560; Gaps 79;

Qy 993 KPAPAP-----PPQNLQPSD-APQPGSSPRGKR 1024
Db 381 KTLPPAPVLMILSTGVLCPFYMINQNGVKSLIKTPERLSLGEKQPKSPGSPPTPTPTS 440
Qy 1025 SPAP-PADEKFAAEAKQLP-----GDP 1047
Db 441 SQAPQKLDASAAAPSLPPSSPADIAFSLIPAGCAPTVSFGSSLSKSAITVTGEP 500
Qy 1048 CWTSGLPFPVPPREVIAKSPHAPDPSAFSYPAPGH---PLPLGLHDTAPVLP----- 1097
Db 501 SYSSG-----SDSKAP-GRGPSTFVFPKXSLAPTP-----AASPAPSAASPS 547
Qy 1098 -----RPTTINPPLIS-----SANGPVLEROICA-----ISQMSVQ 1132
Db 548 FGSSGKPTLESTFPVPSVAPNIAKMSFPPSTSAVKVNLSEKFTAAATSTPVSSQSAP 607
Qy 1133 LHVPSSEHAK-APVGPVTMGLPLPMDPKTLAPSGVQKQQLSPRQAGPPESLGVPTAG 1191
Db 608 PMSPPSSASKPAAAGPLSHPTPLSAPSSVPLKSLTP---SPSGSAGSSSPVPSMQ 664
Qy 1192 ASVLGTAALGSVPGSITKGIPSTRVPSDAITVRSITHGTPADVLTKTITRIIGDS 1251
Db 665 KSP-RITTPAARPSQAASL-QPAVAEKQHQMKSS-----DEVMAG-----IGEE- 709
Qy 1252 PSRLDGRGSDSLPKGHVIEGKKGHVLSYEGMSVTQCKEDG---RSSGCPHETAAP 1307
Db 710 -----IAHFOKELEELKARTSKACQVGTSEEMKMLRTESDILHTFLLE 753
Qy 1308 -KRTYDMEGRGRAISSASIEGLMGRALPRERHSHLKEQHNIIGSTTQGI-PRSYVE 1365
Db 754 IKETTESLHGDIS-SIKTTLRGFAG--VEAREQNERNRDGYLLLYKRPIDPSS--E 808
Qy 1366 AOEVDYLRRAKLLKREGTPPPPPSRDLTEAYVTQALGPLK-----LKPAHEGLVATVKE 1420
Db 809 AAGQETRLKHQYKF-----AVQDVNDVDLDLMDQLEKQKQRHLVVERETLFTTLAN 863
Qy 1421 AGRSIHETPREELRHTPELPLAPRLKE---GSITGCTPLKYDTGASTGSKKHQDVRSIL 1477
Db 864 -NREITNQGRKRLNHLVDSIQQLRLYKQTSIMELSSAVP-----SQSSIHSPDSDESIC 917

Qy 1478 GSPGRTPPPVPLD---VMAADARALERACYEESLSKR---PGTASSGGSIGARGVITYE 1532
Db 918 NALKTTIESHTKSLPKVPKAKLSPMKQQLNRLFLAKRTTPVPRSTAPASLSAFL---- 973
Qy 1533 LKPRQSPULYEDHGAIPRAGHLPRGSPLYTMEPTRLQEGSLSSSKASQDRKLTSTPRE- 1591
Db 974 -----SORYYEDL-----DEVSTSSVSQSLSESDARTSCKDDEAVQAPRIA 1016
Qy 1592 -IAKSPHSTVPERHHPIISPY--EHLR-----GVSGVDLYRSH---IPLAPPTSIPRGI 1641
Db 1017 PVKTRP-SIQPSLPH-AAPFAKSHLVHGSSFGVGTVAITSASKIIPQAGDSTM----- 1070
Qy 1642 PLDAAAAYLLPRHLAPNPTLYPPLRYGYPDTALENRQ-----TIIN 1687
Db 1071 -----AKTVHGAIPSPSHPISAFOQ-----AAALRLRQMAQAPAVNTLTSTLKN 1118
Qy 1688 --DYTSQOMHNHTTAMAQRADMLRGLSPRESSIALNYAGPRGIDLSQVPLVLP 1745
Db 1119 VPQVNVQELKNNPAPTP-----STAMGSSVPSYSTAKTPH-FVLTP 1157
Qy 1746 -----PIRGPTATAM-----DLAVLPTAPQPFSSRHSS 1775
Db 1158 VANQAKQGLINSIKPSGPTPASQSLSGDKASGTAKIETAVTTPSASQFSKPFSS 1217
Qy 1776 PLSPGPTHLTKPTTSSSERERDRDREREREKSLITSTTVEHAPIWRPGETQGS 1835
Db 1218 PSGTGFNRGIITPTSS-----NFTAAQGA---TPSTKSS 1250
Qy 1836 GSSGSSGGGSSSSPASHSHAHQHSPISTQDALQORPSVLHNTGMKGITTA--VESS 1893
Db 1251 QPDASSGGG--SKP-SYEALPESSP-----PS-----GITSASNTTGG 1286
Qy 1894 KPTVLRSSTSSPVPAATFPPTATCPICGTLGVYPTLMEVLLPK----- 1940
Db 1287 EP---AASSRPVAPSGALSTSSKL-----ETPPSLGLPLPSSLAGETLGSFGL 1337
Qy 1941 ---EAPRVARPERPADT-----GHAFLAKPPARGLEP--ASSPSKGE 1980
Db 1338 RVGQADDSRTKFNKASSTSLSTOPTKTSQVPSGFNFAPVYLGNKTEBPVTSATTTSV 1397
Qy 1981 PPLVPPVSGHATITARTPAKNL-----APHNASDPRA--PP 2015
Db 1398 APPAATSTSTAVFGLPVTSAAGSGVLSFGGTSLSAGKTSFSFGSQNTSVTPSPAPP 1457
Qy 2016 ASASDPHREKTSKPSIIGELRLSYGHSYSPEGVEPVPSGSPSLTHDKGLPKHLE 2075
Db 1458 TTAATP-----LPTSFPLTFSGL-----LSATTPSLPMAG--RSTE 1494
Qy 2076 ELDSHLEGEELRPQGPVYKLGGEAHL---EHLRPLPESQSPS-----LQTAGP 2125
Db 1495 EATSSAL-----PEKPGDSEVASAASLLEEQQSAQLPQAPQOTSDSVKKEPVLAQPAVS 1549
Qy 2126 VGHQHVTLAQHISEVITQDTRHHPOQLAPLAPLYSFPG----- 2168
Db 1550 NSGTAASTSLVALSEATPATTPGVDPARTBTPVPASPSVFGQTAVTAALISSAGPVAV 1609
Qy 2169 --ASCPVDLRRPSPDLVLRPPDHGAPARGSPHSECKKSPRPNTKTSVLGGEDGIEPV 2226
Db 1610 ETSTSTPIAS--STTSIVAGPSAEHAAFGITVSGSVFAQPPAASSSSSAFQLTNTMT 1666
Qy 2227 PREGMTEPGHSAAVYPLLYRDEQOT-----EBSRMGSKSPGNTQPPAFSLK 2275
Db 1667 ABSATPFGQVAVASTAPSLF--GQQTGTAATTAATTPQVSSGFSFAPAGTAPGVFGCT 1724
Qy 2276 T-----BSNSAMVSKKQOEINKLNTNRNEPVYNIQPG--TEINMPAITGT 2322
Db 1725 TFGQASVFQGSASSASV-----FSFGQPGSSVAPAGQPA---S 1761
Qy 2323 GMTYRSQAVQEHASGNMGLAELIRKALMGKYDQWEEPPPLANFNPLNATA----- 2375
Db 1762 STPTSTGSGVFAAASST-----SSSSSFSFGQSSSNTTGGUGLFGQSNAPRFGQSPGF 1812
Qy 2376 -----SLPAMPITAADGRSDHTLTSPGCGGKAKVAGRPSRKAQSPAGLSDGRP 2427

QY	514	KSQBEBDEKKEKAEKEEKEPEVENDKEP----	LIKAKTDDTJSGEDNDKEAIVASGKRK	570
Db	409	TRSSVYEHNTNEDDENNASGSDSSDISEGSSSSNEDSEODNDEEDVDEEDVSEBK	468	
QY	571	TANSVGRRK--GRITRSMAN-----EANSSEAIPIPOGAEIASMELNESSRWTEEMETA	623	
Db	469	HEPEBGKASSGONGHRDSNGDKHEDSEKFPQSPSTSH-----HETSHSPKDSXAY	522	
QY	624	KKGLLEHGRMSAIALRWGSKTVSQCKNFYNYKKRONLDEIIOQHKKEKERNARKK	683	
Db	523	Q-----SRSPSL-----NYOSOSPGEFLSEKEIKOE-----FS	552	
QY	664	KKAPAAASEEAPFVVEDEEM-----EASVSGNEEVEEAEALHAS---	727	
Db	553	PTTSASSSDLELDWEMPNDPLTKMLERMWRPFIIVSSFVNRIDEIVELNOKARAIYEK	612	
QY	728	--GNEVPREGSCGPATV-----NSSDTESISPH	755	
Db	613	FTGRFPFR--CNDBEVLISIQKIVHEPRDYIYENPCSELEVRAIRDRKJSDTDADLDR	670	
QY	756	TEAKDTQONGPKRPATYTGADGPPPGPPTPPRRTSRABIEPT--PASEATGAPITPPAP	813	
Db	671	ATDSKEIREDROPAGRTSGR-----PELDESRTNRUSFDSHTNPABELAOSSHLCIGPM	724	
QY	814	SPSAPPVVPRBEKEETAAAPVYEBEGEQPPAAEBLAVDTGAEBPVKSETEBAEG	873	
Db	725	TPSTP-----FPISQPLVUTTH-----	742	
QY	874	PAKGDAEAEATNEGALKAEKEGSGRATTAAGSAGAPODSDSATCSADEVDEABG	933	
Db	743	-----LPGTSQPSISGCIITPRS-----	760	
QY	934	KNRLLSPRSLTPTGDPANASPOKPLDLOKORAAIPIPIQVTKYHEPRRDAPTK	993	
Db	761	-----SQPPLMSFVSRHNSWSSTGRPASITGLRHQSWFPP--DVSLRP-----	803	
QY	994	PAPAPPPPONIQPESDAPPOOPGSSPPRCKSRSPAPPADKEAPAAEAOQLPGDPPCWTSG	1053	
Db	804	--PLPR-----THDEMAAPR--GTPPRSRSSEITWPLASPPFGPIQUL-----LTM	847	
QY	1054	PPPVPRREVITAKSPAPDPSPASVAPRPHPLPLGLHTAPV-----LPRPTISNP	1105	
Db	848	PVPPPHPLIAATSTGTHSVSSASHSTPHSI-----SGTVHCEPSNKSITQOPTKSR	901	
QY	1106	PLPIS-----SAKRPSVLEROIGIISOGMSVOLHVPSYEHAKAPVGVVTGGLPLPMDPK	1159	
Db	902	PEKVDIRDITJSKSPSNAINALQARSSM-----TSG-----DPK	937	
QY	1160	KLAPSGVQOE--QLSPRGOAGPESIG--VPTAOASVLRGTALGSVGGSIYKIP	1213	
Db	938	KSAPSTPVLRDAGSGLVAQIMSNCP--NGLRKLPIREKSSALQNIQNHQPPHMANSTP	996	
QY	1214	STRVSDSAIYTRGSIHTGTRADVLYKGTITRIIGEDS-----PSGLDGRDLSLPG	1266	
Db	997	ST--PSTS-----TH-----QANFQKXEXEKKKKEKEKEEREAREAREKREKTEER	1042	
QY	1267	HVITYEGKGHVLSYEGMSVTQCSKE--DGRSSSGCPHETAAAPKTYUDMEERVGRALISA	1325	
Db	1043	NKRKEMERAKRLIEDRQERKRKKERBERKKEKVKAKYKAEKLLKKKKHKKDSSDES	1102	
QY	1326	STEGLMGPAIIPPERHSPHHLKEQ--HHI-----RGSITQIP--RSYVEAOEYLRREA	1375	
Db	1103	DSDSNDELIDLVRKSTKEMTQOEKHOALALLSKGIIENLKSRRRSOKRAHDSFEKQOQ	1162	
QY	1376	KLTKR-----EG-----TPPSPRSRLTAUYTOALGP	1404	
Db	1163	KSQGRVVLIESDDGCKDGKNSNBEESDEKADLPPPPAPS--LSHSAQR-----	1216	
QY	1405	LK-LKPAHEGLVATV-----KEAGRSIHE-----I	1428	
Db	1217	LKVLKERKEGELJYTSDDDEHNDAG--EIHQORLTEDRENRRKOKSLTAYSDDEOGERNAV	1275	

QY	1429	FREBLRHTPELPLAPR---	LKESG1TQGPRLKTDYTA5TTSK-	-KHVRS1LIGSFORT	1483	
Db	1276	PKMRKRDDBSEAAAKHGMWAKDO	KOKRKKTLERRSSEDESCKNAKRD	FRD1-----	1328	
QY	1484	FPYHPLDVMADALRACAYESL	KSRPGPASSGGG51AR-----	-G	1525	
Db	1329	-----PHEVDVSDEETE---	DGSRKRQSTST51SNVTKERKESG	KTPRLRVPEPTG	1379	
QY	1526	APV1VEIKGRQSPRLTYEDHGA	FAFAGHLPRGSPVTKREBTPRL---	OEG5LSSXASQ	1581	
Db	1380	TP1LSPK1LSPK-----	HL-----SPK1TSS1TKSS1SDHNL	1LSPQRNR	1421	
QY	1582	DRK1TSPRE1AKSPHSTVDENH	HP1SPYENH1RGVGDYLRSH1PLA	PDPS1PGR1	1641	
Db	1422	TTSSTSTATSSKHEAL51PE---	KPLSPVTKASSV51D-----	DP-SIRDEF	1467	
QY	1642	PLDAAAYLYLPH1LAPRPT---	YPHLYLPRYL1RGPRDPAALEN	QTI1NDV1T5QOMHNT	1699	
Db	1468	SMNSAAD-----	SPWSTGRFVNLTKAMKANSTPR	KKETDQAVQ51PDEE	1515	
QY	1700	ATAAQAQADM1LGRSPRESS1	ALNYAAGPRG1DLSQVPHLYL	VP1PRTPGPATAMDRLA	1759	
Db	1516	ADEPRQYPRD---G1STNKEVS	-----GQDPH---N1K1T-----	1545		
QY	1760	YLTPAPQFSSRHSSSP1SPGCP	PH1LT-KPT1TSSSERERDRD--	-REDRDRER	1810	
Db	1546	-----EPLNNGHTD1LF	SPSSAHASEKQSTKSEDDME	ED5ELVMEKEVMEQV1AOE	1599	
QY	1811	-----EKSL1TSTTV-----	-EAP1WR1GTQD5SGSSSGG	SGGSSSRPA	1852	
Db	1600	VHVSEPSRMEEEKLETS	VPVKKEP1KMEBSPQOTPR	1L1SN5QDTRG-----A	1652	
QY	1853	SHSHAHQ-----	SPI---SPRTDQALQO---RPS	VLN1NTGKIG1ITVABESK	PVL1ASTS	1902
Db	1653	VANH1HEHNDVAQTR1Q	Q1RASQ1VQABRPRVA	VD5QONGV1V5QOSQ	BRPM5SQ	1712
QY	1903	-----TSSVPRPAAT---	FRP1TNCPSGLTDYVUTL-----	-MEV1LLP	1939	
Db	1713	SDMAQNL1LSSKQD	IND1LAALKHKNRBEALQ	ATRQDSCG1FQNL1LHAQ	NGQNMTPMLQ	1772
QY	1940	-----	KEARV1AR-----	1947		
Db	1773	LKAAFAQOQENEAQNMQAK	KQOTINKDRIKQEVVKRMV	EENERRKVEBDRREKORKE	1832	
QY	1948	PERAPR---ADTGHAF1AK	PPARSGLEPASS--PSKQ-----	-SEPRPL---VPV5GHA	1992	
Db	1833	BERQRLAAATTAATMA1Q	KAELKQAEVPRHGFQV1L	SM1TPEAS1LYE0FGGLSSY1	1892	
QY	1993	TIARTPAKN1LAPR-----	-HASDP1RAPPASDPR	HEK1TQSKRF51QIELELR	2039	
Db	1893	NRDS1GANTGV1LH1PTQ	S1QRPSSTASTSSNP	FRAP1QPSASV1QNT--	IDPAB1E1RVQ	1951
QY	2040	SLGH1GSSYSPEVGEVPS	PVSS-----PSLTHDKL	PKH1LELDK-----	2079	
Db	1952	RWFYKPLTMSABEATW	AAVASSDPNPRPA1STVD	LA1MLQDLQAAQAAQV	PVVTTA	2011
QY	2080	-----SH1LBE1LRKQ	RGPKVLGGBAHLPH1LR	LPESQSSP1LY1Q1AR	GVGHQV	2133
Db	2012	STPNR1L5LE1TL1TAS1	LAN1ATGGA-----LNP1	LMAL1TSS-LM0QSP	VYOG1ARVL	2064
QY	2134	-----TLAQH-1SEV	1TQDYTHHNRQDLSAP	RLP1RPSFGASCVL	DLRRP5DYL	2185
Db	2065	LTMNMG1L1AHQ1SSELLA	-----TMNQOET1MAL1AA	-----RNG1LPRAM	2105	
QY	2186	PPPDH--GAPRAGS-----	PHSEGGKSPENPKTS	VLGGGEG1EPVSP	BBEGMTBRGH	2236
Db	2106	FOONQOPMPRQOGG	FA1PTVLPHN-S1KRNKADOL5--	VGGVSD-----R	2147	
QY	2237	SRSAYPL1LYDNGQ---	-TERPSMG5KSPNTSQ	PPAFKSL1TESNSAMV	KSKQO1E1NK	2292
Db	2148	KKS1CPLAHMTQGGQ	QPPRQOPQAVAPRPS	SP-----RKSMPEN1	PREMKKE	2198
QY	2293	KL1N1TRNEPEYNI	5QDTE1FNMPRA1TG1G	1M1TYSQAVQENAST	TMG1EAL1RKALMG	2352

Oy		MEALKEKVERI-----ENNPRRRAKESKVREYVEKOFPEIRRKORELOERNOQSVGORG	358
Db		585 LKSLEKENKAIKOCGSSETSRRRGROLPTPKEDTQTLO-SLKENOELMKLSLEGULE	643
Oy		359 SGLMSAARBHEHVESLIIDGLSOGENLEKOMROLAVIPMLVDADQRIKFIFNMGLMAD	418
Db		644 TFL-PPGTENOELVSSIOENLESITALEKENOPPLASPEV--GBEALRPIITXEN--QE	697
Oy		419 PMKYVKDROVANMMWSEOEKETFR--EKFMHPKNFGILASFLEKTVAECVLYYLTTKN	476
Db		698 PLRLIED-----ENKEAFRSLEXENQP-----LKTLIEEDOSIVAPLEAT	737
Oy		477 ENYKSL-----VRSTRRKCSQQQQQQQQQQQQQQQQQQPMRPSQOEKDEKEXKEAE	530
Db		738 ENHKSLSLEBODETTRTLEKETYOQRSLGEODOTTLRP-----PEKDIPLEKSLD	791
Oy		531 KEEEKPEVENKEDIKEDTDTSGEUNDEKEAVASKGRKTANSOGRRKGRIITRSMANEA	590
Db		792 OEIARP-LENENOEFLLSIKES-----VEANKSLETETILBS-----LKSAGE	834
Oy		591 NSEEAITPOQSAGIELASMELNESSWTBEEM-----ETAKG-----LLEH	630
Db		835 NLETKSPETOAPL-----WTBEIKKSGNSSRKSGNSRTTGCGSEPRDIQTP	884
Oy		631 GNMFAIKRMVGSKTVSOC-----KNFYNYKKRONLIDEIIJOHLKMKERNARRK	683
Db		886 GRGESGITIEIGSNHEPGFEISRGVDKSCONLEENLNKGEVOESL-----RSLEERG	939
Oy		684 KKAAPAASEBAAPPVVEDEEME-----ASGVSGNEEWEVEAEALHASGN--VPREE	735
Db		940 QELPOSADVORWEDIVEKDOELAOESPGMAGVNKOKEALINLBODGFCKEVEBOQG	999
Oy		736 CGGPATVNNSTESIISP--HTEAADTCQNCPKPATL-GADBPPEGPPTPPRKS	790
Db		1000 LNA-----TEEWFPGBSHPNPNPPKORGLVEGASVKGABELO-----	1039
Oy		791 RAPIETPASSEATGAFTPPRAPPSBAPPVVPYEKEEBEETAAPPYEEBEOKPPAAEF	850
Db		1040 ----DBEQSQOVTFGLQAPQGLPEALIEPIVEDIVAPEGDOSPEWMLOSE----PAMGE	1092
Oy		851 LAV-----DTGK-----AEPVYSRCST-----BAEBSPAKDAEALAAT	886
Db		1093 SAAGAEPGLGGVGLGDPRGHILTREEVMBPLEEBSLEAKVQGLEBR-KOLEEA---	1147
Oy		887 AEGALKAEKKE-CGSGRAATTAKSAGAPODSDSATCSADEVDEAE-----GDKRNLLSP	940
Db		1148 --CGJGTERSELPGKSRDPWPBPRGREGSEAEAPRAEAPETUIGHNGSD-----AP	1200
Oy		941 RBGLLTPFGDDPRANASPOKPELDLKKOLKRAAIPIQY--TKNHPPREDADAFTKPAPA	998
Db		1201 SP---WPLSEEEBED-----VPPVALVSPEPTYTPILED-----	1232
Oy		999 PPPPONULOPESDAOPOPSSPRGKSRPAPPADKCAFAAEAOKLPGBPCTWSGLPPVP	1058
Db		1233 ----PELOQAOBESQCLASMGVGRABAGVSESQELDS-----GBIP-----EGL---	1274
Oy		1059 PREVIKASPHAPDPSAFSTAPFCHPLPLGLHNTARPVLPR-PPTISNPPLISSAKH---	1114
Db		1275 QEEGESSRESEDEDELGETLPDSTPLGFLYRS--PTSFRMTPLRESRHPLKETGKEGWD	1331
Oy		1115 PSVL-----EROIGALSQMSVOLHVPSY-HAKAPVGPVTMGLPLMDPKXL	1161
Db		1332 PAVLASBGLPEPBSEKBERGEBEGCGRDSDLBBEPDLGTAEAF-----LGVPGEV	1383
Oy		1162 A-PPSGVKGQBQLSP-----RGQAGPPBSLC--	1185
Db		1384 ABPLGQYQOLLIPAAWDRDGESDGFADSEESGEBEDEOEBRGPGAARGWGPGSSVSL	1443
Oy		1186 --VPTQOBAVLACTILGC-SVP-----GGSITKGIPSTRVPSDS--AITYGSLITHGPA	1235
Db		1444 QAUSSSGGRGEFLSDSVSVPMDDSLRKAVA-GAPKTLBFHSQDSNAPSGBSEESDPV	1502

```

QY      1236 DVLVYGTITRILIGEDSPSLDRGRDLSFKGHVIT-----YEGKKGHVLSYEGG-MSV 1286
Db      1503 SLERDKV-----PGPLTIPSGMEDAGADITIGVGNQPNLEGSQHV--NGVWNG 1553
QY      1287 TQCSNEDG-----RSSGSPHETAPAKRTYDMMEGRVGRALSSASIEGIMGRAIP 1336
Db      1554 LEQSSSESGARNALVSEGDGSPPEEGEGSLKR-----SSAG----- 1590
QY      1337 PERHSPHLKEQHIRGSIYTGIPRSYVENQ 1368
Db      1591 ----APVHLGQGGQFLKFTQREGDRESWSGED 1618

RESULT 51
T20531
Hypothetical protein F07A11.6a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20531, T27776
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: T20531
A:Accession: T20531
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2526 <MTL>
A:Cross-references: EMBL:Z66511; PIDN:CAB54210.1; GSPDB:GN00020; CESP:F07A11.6a
A:Experimental source: clone F07A11
R:Gajdstry, S.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20417
A:Accession: T27776
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2526 <MT2>
A:Cross-references: EMBL:Z69904; PIDN:CAB54501.1; GSPDB:GN00020; CESP:F07A11.6a
A:Experimental source: clone ZK20
C:Genetics:
A:Gene: CESP:F07A11.6a
A:Map position: 2
A:Insertions: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3; 1

Query Match      2.7%; Score 356.5; DB 2; Length 2526;
Best Local Similarity 18.0%; Pred. No. 3.5e-05;
Matches 490; Conservative 338; Mismatches 972; Indels 927; Gaps 123;

QY      155 HTPELELVPRLLKSELIQNMDRVDRREITWEEQISKLKKQOOLEE-----AAKPE 209
Db      47 HGNPEILVSTMARGAL-----PARSTKPEEDSIEYKISAYIBE 87
QY      210 PEKEVSPPIESKRSLSVQIITYDENRKKAEANRLLEGIGPOVELPLYNQPSDTRQYHN 269
Db      88 PTQN-SPIP-EFPESTACVVE-----IQSGTPE-----RDLFEL 122
QY      270 IKINQMAEKLLIFKRNHARKKQKQCFQRYDLMELAEKVERIENNPRRAKESKY 329
Db      123 VK-RHKSRSQVPIDIQLESTTEPGMKKARVNHRYRLDTDSLKADKSLILGR-----PKF 175
QY      330 REYY-----EKOPEIRK-----QRELQERMQRVQGRGS 359
Db      176 RYVYPTSGQKHPOCHPSTSYAIPKLKGPHLLKASCVYVPHLDHRSPOHYRRFESYQ 235
QY      360 GLSMSAARSHHVSIIIDGLSQENLEKQMRQLAVIPMLY-----DADQ 404
Db      236 VIDVMVSNDKAAVAVQFTNIDDAQKALQDTNIPKPSYSQSRPSHRHIIIFLPIECTN 295
QY      405 QRIKFI--NMGLMADPMKVVYKQVNMNMMSQEKETFEKEMQ--HPKNFGLIASFLER 460
Db      296 EIMILITSLSRIVIDICVDMWDRSAVITLDMIEPANILLKRWKLIVGRNNGF-----EH 349
QY      461 KTVAE-C--VLYYLTIKKNENYKSLVRSYRRRGSGQQQ--QQQQQQQQQQQQQQPMP 513
Db      350 KYAVDFCGDRFNLVYINRKENIE--VAARSSSFTSKSENDQSSSFSSSSDRQNLHPDQ 408

```

```

Db 777 TSGSTQTPGSSVSTSAAILTSTQGSVSTNPGSTVTRPSTVSSGSSGTVTVGSTRAS 836
Qy 1104 -NPPPLISSAKHPVLERQIGAIISQKMSVOLHVPYSEHAKAPVGVMTGLPLMDP--- 1138
Db 837 TSGSSVASSPAPSTSQPNPSTSGSSSMITQSPFPGSTSPVSEST--TSPSGSPGTTL 894
Qy 1159 -----KKLAPSGVKEQJLSPRQAPPPESLGVPTNOEASVLRGTALG 1201
Db 895 TSTSPSPQSTTTGSTQGSTPGISTSEMTSQSTQTPGSTGTVTTPSTSDSTSSG 954
Qy 1202 SVPGGSITKG---IPSTRVPSDAITYRGSITHTG-----PADVLYKG----- 1241
Db 955 STTVVSGTEGSSSPSTPSQNTNPTSGSSSMSTQTPSSQSTSPVSESTSGATSSSGSP 1014
Qy 1242 -TTRTIGEDSPSLDRGREDG-----LPKHVITYEGKKGHVL-----SYEGG- 1283
Db 1015 GTTLTISIPSPSPSSSTTSSQSTSPVSTISQGSTETPGSTGTVTPSTVSGSASGS 1074
Qy 1284 -MSVTCQSKEDRSGSP-PHETAPKRYTDMMEGVRAISASIEGLMGRAPPER 1339
Db 1075 TATMGSTIASSTSGSSSTSPNSQSTSPS-----TSGATSSPGSSGTTLTISP-- 1123
Qy 1340 HSPHLKEQHHRIGSITQGIPRSVEAOEDYLREAKLKEGTGPPPPSRDLTEAYKT 1399
Db 1124 -SP-----SQSSTIGS-SQG-----STSPVETTSQDMTSSQST 1155
Qy 1400 QALGFLKLPKPAHEGLVATVKEAG---RSIHEIPREELRHTPELPLAPPL----- 1446
Db 1156 QI-----PGSTGTVTQTPSTGSGSTGSGTSGEITSGSTQTPRSSLSSTPASTSTQOSV 1208
Qy 1447 -KEGS-ITQCTPLAKYDT-----GASITG-----SKKDVSLSISP--- 1480
Db 1209 STNSPSTVTPQSTVTRSGSTSGSTVTTGTEGSSSTSGSSSATSSSSPSTQSPSPS 1268
Qy 1481 -GRTPPVHPLDVADARALERACYERSLKSPCTASSGSGSIARGAPVIVPELQKPRQ 1538
Db 1269 TSGSSTPPTNP-----SQSTSPVSTTGTGMTSHGSTQTPSTIGSTVYT 1311
Qy 1539 SPLTYEDHGAPPAHLPRGSPVTMEPRPRLQESGLSSSKASQDRKLTSTPREIAKSPHS 1598
Db 1312 QPST-----VSGSNSSGSTVITI-----GSSEASTSGSSPK--TSPSSI-----S 1348
Qy 1559 TVPEHHHPHISYEHLRG-----VSGVDLYRSHIPLAFDPISIRGIPLDA----- 1645
Db 1349 PVPSTSPPTSTTPASTSGSTISDVSVST-TSLAPLS---SSLSTVPSSTQSSSTSE 1404
Qy 1646 -AAATYLPRLHAPNTYPLHVPYLLRGYDPAALENROTINDVYTSQOMHNTAT 1701
Db 1405 GSSKASSSPVPGQSTSTPNP-----TGSTHS-STLSSSTSGSTQH-----T 1446
Qy 1702 AMAQADMLRGLSPRESSIALNYAAGPRGIIIDLSQVPHLPVLVPTPGTPATA-MDRLAY 1760
Db 1447 TMSK-----ASSGSTSPSTNSQTG-----STV-----TMGSSSTSGVSTSGA 1483
Qy 1761 LPTAQPPSSRHS-----SPLSPGSPTHLTKTSTSSSERERDRDRERDRERHS 1813
Db 1484 SSTQOMSTQSSSGSTVASTASBPASSTAPSSSTGTMSS----- 1524
Qy 1814 ILTSTTVYHAPITMPGTEQSSGSSGSGSSSRPASHAHQHPISPRTDALQO 1873
Db 1525 -TSGTGSTISSESTTASASQSTGYTMGSSSTSGVSTSSASTQOMSTQSSSAG 1582
Qy 1874 RPSVLHNTMGKITTAVERBSKPTVLRSSTSSPVRPATFPATHCPLGGLIDGYPTLM 1933
Db 1583 STVASTAGL--VSTSTVPSSTGTGWTSGT-----VGSTI----- 1617
Qy 1934 BPSVLLPKAPRYARERPRADTGHAFIAPKPARSGL-----BPASSPGSEPRPLV 1985
Db 1618 -----SESSTTASAS--SQTGSTVTMGSSSTSGVSTSSASTQOMSTQSSG----- 1662
Qy 1986 PVSCHATTIARTPAKMLAPHNHASPDPAPPASASPHEREKTSKPFSIOELRLSLGYNG 2045
Db 1663 -SAGSTVA-----SSTTGIVSTSTVPSSTGTGWTSGTGVG 1697

```

```

Qy 2046 SSYSPEGEVPEVPSVSPSLTTHDKGLPKHLELDKSHLEGELRPQPGPVKLGEAAHLPH 2105
Db 1698 STIS-----ESSTASASQCT-----GSTVTMGSSSTS--- 1725
Qy 2106 LRPLPESQPSSSPLLQTPAGVKHQRVVTLAQHISEVITDQYTHHPQOLGAPLPAPLYS 2165
Db 1726 -GVSTSSASQCPQMSTSSQSSAGSTVSS----- 1753
Qy 2166 PPGASCPLYDLRPPSDLYLPPPHG-APARGSPHSBGCKSPPEBNKTSVIGGEDGIEP 2224
Db 1754 -STASPAASSTASSTGTMTSGSTSGGTGMSQSTASSTTSHGTSVTLGSSSTSSNQ 1811
Qy 2225 VSPPEGMTEPHSSSAVYPLLYRGEQREPRMSSKSPGNTSQPPAPFSKLTESNAWVK 2284
Db 1812 MSTQSGSVSGSTVASTAGLVSTSTVPSSTGTGMSGTVG-----STISESTTASA 1865
Qy 2285 SKQEIINKLNTNHRNEPEVNIISQPTGEIFNMIPAITGTGLTYRSQAOVHAETMGLBA 2344
Db 1866 S-----SQTGSTV-TMGSSSTGVSSTSSASSTQPMSTSGQ--- 1900
Qy 2345 IIRKALMGKYDQWESPELSANATNPUNASASLPAPMTAADGRSDHTLTSPGGG--GK 2402
Db 1901 -----SSAGSTVASTAGLVSTSTVPSSTGTGMSGTVGSTISSE 1941
Qy 2403 AKVSGRPSRKAQKAPAPGLAGDRPPSVSVHSEGDCKRRPILTRVWEDPSS-AGSTP 2461
Db 1942 SSTASTSTQSTGTVTGTSTGTGNTSPSPRSLSQ---ITTPSPSQSTESTQTSLPSSSP 1997
Qy 2462 PPYNPLIMRLOAGVNAS 2478
Db 1998 SPSTHVSASSSGTMTSS 2014

RESULT 50
S21424
nestin - human
C:Species: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
C:Accession: S21424
C:Jaccard: 0.7; McKay, R.D.G.; Zimmerman, L.B.; Lendahl, U.
submitted to the EMBL Data Library, May 1992
A:Description: Characterization of the human nestin gene reveals a close evolutionary re
A:Reference number: S21424
A:Accession: S21424
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1618 <DAH>
A:Cross-references: EMBL:X65964; NID:g35018; PIDN:CAA46780.1; PID:g35019

Query Match 2.7%; Score 358; DB 2; Length 1618;
Best Local Similarity 21.1%; Pred. No. 1.9e-05;
Matches 323; Conservative 196; Mismatches 511; Indels 502; Gaps 77;

Qy 97 KSEMEFISKAPRL-ELLP--DPLRPSPL---LATQPA--GSEDTLKRSITGLKEP 147
Db 339 KLELOFPPTPEGRKRGSLPVLSPSLPSLPATLETVPAPLQKQBFLOARTPLASTP 388
Qy 148 VSPSPRPTDLELVLPRLSKELLQMDRVDRBITWVEQOISKL-----KKKQOOLE 201
Db 369 I-PPTPQAPSP-----AVDAEIRADQADPLSLLOTQGRKQAPPELR 428
Qy 202 EEA-AKP-----PEPEKP-----VSP-----PRIESK----- 223
Db 429 AEARVALIPASVLPGEPEGQROEASTQSPEDHASLAPLSPDSSLEAKDSESGSRV 488
Qy 224 -----RSIVQIYDE---NRKKAEAAH---RILEGL 248
Db 489 PSICGEBEGQIWLVEKETAIEGVNSSLQOEIWEEDDAKKELODQVPLEKXTIKSL 548
Qy 249 GPQVE--LPIYNQSDTRQYHENIKINQAMREKKLILYKRRNHARKQWKQKFCQRYDOL 305
Db 549 GBEIOESIKLENQSHETLE-RENOECPRSLIEDL-----ET 584

```



```

QY 1059 PREVIKASPHAPDPSAFSAPRPGHPLPLGLHDTARVLPREPPTISNPPPLISSAKHPVL 1118
DB 1447 PPVAKSS-----VSSIDDEIR 1464
QY 1119 ERQIGALISQGMVOLHVPSEHAKAPVGPVTMGLPLPMDPKKLAPFSQVKOBQLSPRGOA 1178
DB 1465 DE-----FSMNSADSPMSITGRMVLTKAMKAFNST-----1497
QY 1179 GPES-----LGPVTAQASVLRGALSGVPGGSITTKGPSTRVPSDAIITYRG 1227
DB 1498 -PPKVSYSLLIIDCYMLGMMAKNSSGQHDSGSSSDSGTSSDSDSDDEVPQOT 1556
QY 1228 SITGCTPADVLKGTITRIIGEDSPRLDRGREDLPGHVIYEGKGHVLSYEGMSVT 1287
DB 1557 EPTVTSIPVADNSGSPENVAVE-TPIVSQTPRE-PEPFTISE-----1597
QY 1288 QCSKEDGRSSGPPHETAPAKRTYDMEGVRGAISSISGLMGRALPPERHSPHLKE 1347
DB 1598 -----QSSSEPR-AVPE-----CPASVPEQMETSGNVEPVSSEHDS 1635
QY 1348 QHHRGSIITQGI PRSIVEAQEDYLREAKL-----KREGTPPPPPSR 1391
DB 1636 HEHGSEVAVESQOQPLEHSEKELENNKILDAEHHBEQVQDESVESIPAPSDER 1695
QY 1392 D-LTEAVTQALGPLKIKRAHEGLVATYKAGRSIHEI PRELRTPLPLAPRLKGS 1450
DB 1696 DPTVQAE-----KSAH-TLISQETDOAQVSIPIDEE-EADEPQY-----1735
QY 1451 ITGQPLKYDTGAST-----TGSKGDVRSILGSGRTPPPHPL-----DVMADARALE 1500
DB 1736 -----DGIISTNEKEVSKDPHNK-----PREPLNNGHTDLTFPSSSA 1775
QY 1501 RACEESLKSRRGTASSGSGIARGAPVIVPELGRPSPLTYEDH---GAP-----1550
DB 1776 HASEKOSTKSEDDMEDS-----ELVMEKEVMEQVIAQEVHVSEPSMEEEVYL 1827
QY 1551 -AGHLPRESSPYTMRE-----PPRLQEGSLSSKASOD-----RUTLSTPREI 1592
DB 1828 ETSVPKKEPIKMESSPEOTPR-----DLISNNESQDTPGAVNNHLENHDAVOPTIQL 1882
QY 1593 AKSPHSTVPEHHPHISPEYHLLRGVSGVDLYRSHIPLAFDPTISPRGIPLDAAAYLP 1652
DB 1883 QPASQHOVAQSPRAVAPDQNGPVLVSQOSQPSMSSQSDMAQMLILSSKOINDLA 1942
QY 1653 RHLANP-----TYPLYPYLLRGY---PDTALE-----1680
DB 1943 AKLHNPRPALAOATRGDCSGIFQHLHLHAQNGQNMTEPMLQLKAAFAQOQENANQM 2002
QY 1681 -----NRQTIINDYITSQ-----MHNTATAMAQADMLRGSPRESSIALVYA-----1726
DB 2003 QAKMKQOTINKRIKEQERVKMYEENKVEEDREKQKEERQRLAATAATMATQ 2062
QY 1727 -GPRGIIIDLSQVPH-----LPVLVPT-----GTPA-----TAMDRLAVLP-----1763
DB 2063 KAAELKQKQVPRGRGFOVLISMTPERSLYEQFGGLSYINRDSIGATNGVHLPIFOS 2122
QY 1764 --APQFSSRHSSSPLSPGPHLTKPTTSSSERDR-----DREDRDREREKSIIT 1816
DB 2123 IORPSTASTSNPKAPLPQPSASVNONITIDPAIEIIRVQRFYKPLTMSAEKATVA 2182
QY 1817 -----STTVHAPIRWRCPTQSSSSSSSGGSSSRPASHSHAHQSP-----SP 1865
DB 2183 VASSDPNPATISTVLAAMLQ-----QLOAQAOAAQOVVYTTA 2223
QY 1866 RTODALQORPSVLAHNTGMKGIIT-AVEBSKPTVLST-STSSVPRAATPPPATCPGL 1922
DB 2224 STPNPLSNITLSTASLANLATGALNPLSMALITLSLNOSSP-----2267
QY 1923 GTLDGYPTLMEPVLLPKAPVAPARPER-----RADTGAFLAKKPPARSGLEBASPP 1975
DB 2268 -----VYOGIARVLLTMNMGOMLATIHQTELLATNMQOETLMAALLA---ARRGL-PPAMP 2318
QY 1976 SKGSEPR-----LVPPVSGHATTAAR-----TPAKNLAAPHNA-----SPDP- 2011

```

```

DB 2319 QONQPOWPAQGFALPTVLPHMSLKNNAKQQLSVGVSDRKSCLHAMIQGGQPPPP 2378
QY 2012 -----PAPPASADPHREKTOSKPSIOE-----LELSLGHSSSY 2048
DB 2379 QQPMQAVAPAPRPPSPRPKMPENLPMEKNEKMERKEILRLDILLLELG-----2432
QY 2049 SPEGEVPSVPSSSLTHDKLPHKLELDKSHLE-----2083
DB 2433 -AEDEBD---QKFDL---KQPTSEBDTDSKADSMKAEASAPRILSRSTGMNNGS 2484
QY 2084 -----GELRPKQPGCVKLGEAHLPHLPLRESQPSSSPLLQ---TAPGVKHQVVTIAQ 2137
DB 2485 PSAGTTSPTSSSSISSG-----PDSPPL-EGEPLNGEFMDMLTEVAQKREQENTDAL 2537
QY 2138 HISEYITQDYTRHNP 2152
DB 2538 SAKIVDEQSFQHP 2552

```

RESULT 48

T38236
 hypothetical protein SPAC23A1.17 - fission yeast (Schizosaccharomyces pombe)
 C1:Species: Schizosaccharomyces pombe
 C2:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C3:Accession: T38236
 R1:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, February 1998
 A1:Accession: T38236
 A1:Status: preliminary; translated from GB/EMBL/DBJ
 A1:Molecule type: DNA
 A1:Residues: 1-1611 <MUR>
 A1:Cross-references: EMBL:AL021813; PIDN:CA16991.1; GSPDB:GN00066; SPDB:SPAC23A1.17
 A1:Experimental source: strain 972h-; cosmid c23A1
 C1:Genetics:
 A1:Gene: SPDB:SPAC23A1.17
 A1:Map position: 1

```

Query Match 2.8%; Score 367; DB 2; Length 1611;
Best Local Similarity 19.7%; Pred. No. 9.2e-06;
Matches 347; Conservative 201; Mismatches 698; Indels 514; Gaps 77;

QY 23 SLSPVQARHTHTVGLLEQYHNSRDVASHSPSGIIPQRRRPSELSPQP-----GNE 77
DB 94 SLGVAPSTRSTKSTIS-NNTTEHLAGDTESVLSANDPISTSEALTKIVDVTLSPAFGND 152
QY 78 RSQELHLPRESHSYLPGLKSEMEF-IESKPRLELPPPLRPSPPLATGQAGSEDLT 136
DB 153 SNS-----RPHSLKXNVEKHTSAPYTISEPRPSCSTENDSL---PLRATHITTGEDAA 204
QY 137 KDRSLTG-----KLEPVSPSP-----PHDPELELVPPRLSKELLQNDNR 178
DB 205 TGAATVNTTHTTHTTSTVTSFVIPGNPNSVFLVDCTHSCCPDPL-PIATTHGHSRLYDN 263
QY 179 VDRE-ITWVEQISLTKKKQGLEEAKKPREPKVSPPTIESGHSRLVQVIYDENRKK 237
DB 264 ASASATYLER-----THPAAS-----STMATESHQPSPA---DSQAE 300
QY 238 AEAARILLEGCPQVELPLYNQ---PSDTRYHENIKINQAMRKLLIYFKRNARKQW 294
DB 301 LSKQGVAKDDDPFVSNVTANSDDEPASSKRAKPLTLDNRARSQGLNLDPOKRGKSGEHI 360
QY 295 KQKFCQRYDQMLEALEKVERIENNRRAKESKTYREYKQPPETIKORELQEMQSKV 354
DB 361 SEQEDDEYDD-----AESDEHMSYSTHEPE-----386
QY 355 GQKSGGLNSARSHVESEITDGLSEQENLEKQKROLAVIPMLYDADQOKIFLNNWG 414
DB 387 -----SEPEDDESEKDD---ENMDVEEBQO-----EQEBEQI-----418
QY 415 LMADPMKYKQDQVNMNMSEQEKETFRKFMQHPKNGFL-----IASFLRKTVABCV 467

```

QY	1546	HGAPRAGHLPRGSPRYTMEBPRRLOEGSLSSKASQDKLITSTREIAK--SPHSIVREBH	1604
Db	646	ATPP--GKKP--GSPPSPFRITGPPGYG-----TSPRPGSTFKPS--PRVG	687
QY	1605	PHISPYEHLKGVSGVDLYRSHIPLAFDPDISIPRGILDAAMAAUYLRHIAEMPTGYHL	1664
Db	688	FGPL-----PRAGFGLP-----SLPPPAAMAS	711
QY	1665	YPPYILIRGYPTALAEHQIILINDYITSOQHNTATYMAQRADMLRGLSPRESSLALNY	1724
Db	712	GP-----LSAQIKQEPAAEYET-----	730
QY	1725	AAGPRGILDSQVPHLPVLPPTGTPATAMDRLAYLPTAROPFS--RHSSSPSPGPT	1783
Db	731	---PSPVPPARSPRP---PKVDPVREHAQSANFNGHLDRGPRSCASDLUYFPLEGS	784
QY	1784	HLTK-----PTTSSSERDRDRERDRERBKSLITSTTVE--HADIW	1827
Db	785	KLAKRADLVEKVRREARERAREREKEREREREKERERELERESVUKLOEGRAPVE	844
QY	1828	RPGTEQSSGGSSGGGGSSSRPASHNAHQHSISRTODALQORSVLHNTGMKII	1887
Db	845	CP-----SLGPVPH---RPPFEQS--AAVTPPYL--GPDTPALR	878
QY	1888	TAVESEKTVLRSTSTSSPVPRPAATFPRAHCPLG---GTLDGYUTIMEVLLPKCAP	1943
Db	879	TLSEYARBHVMSPGNRBNHP-----VYPLGADVPGLLGYVVPALYSSDPARERE	928
QY	1944	RVARPERRATGHAFLAKPRASGLERPASS--PSKSGPRPLVPVSGHATIAATPAKNL	2002
Db	929	REABERDRLKRGGEFVKP---SELEPLHGVPGGLP---PRPHGGLAQOPGGLH	982
QY	2003	APHNASDP-----PAPRASADPRHEKTSKPFISQELERLSLYHSGSSYPBEVE	2054
Db	983	FPFPHSGPLERERRLAALAPALRP-----MSYAE--RLAERQHAEVVALGND	1031
QY	2055	FVSPVSSSLTHDKGLPKHLEELDKSHLEGELRKQRPQPVYLGGEAHLPHLRPLPSQP	2114
Db	1032	PLARIQMLNVT-----PQHIOH--SHISHHLHOODAHNAASAVH--PLIDL-----	1077
QY	2115	SSSPLLTAPFVKGHQVVTLLAQ-----HISEYITQDYTRHH-----POOLCAP	2158
Db	1078	ASGSHLTRITRPAG-----TLPNPLPHPLHENVL-----RHQLFAARYDLRASILAP	1127
QY	2159	LPAPLYSEFGASCVLRLRPPSD---LYLPRPDGA--PAKSPHSEGGKSPRP	2209
Db	1128	MSA--AHQIQAMQASASAEIQRALTEQQQWIAHNNPLHSAVPLPAQEDYVSHLKEEKSDK	1183

```

RESULT 47
E88320
protein F07A11.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88320
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.muscl.edu/gsc/c_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88320
A:Status: preliminary
A:Molecule type: DNA
A:Releashes: 1-7738 <STO>
A:Cross-references: GB:chr_II, PIDN:CAA93781.1, PID:G3881547, GSPDB:GN00020, CESP:F07A11.6
C:Genetics:
A:Gene: F07A11.6
A:Map position: 2

```

Query Match	2.8%;	Score 368.5;	DB 2;	Length 2738;
Best Local Similarity	18.1%;	Pred. No. 1.4e-05;		
Matches 436;	Conservative 211;	Mismatches 885;	Indels 803;	Gaps 94;

OY	44	AHSHDYASHLSBPGSLIQORRRPRLS----	-EPQPN-----RSQELHAPESHSTYP	93
Dd	635	HEPRDYIYYENPCSELVIRIMWRKLSLTDADDDRADTDSKEIGDQPAGCTSGRPSLD		694
OY	94	ELGKEMEFIESKR-----RLLELPDLRLSPPLATGQPAGSBDLTDRSLGTGLCEPV		148
Dd	655	ESRTNRSLFSDTHHBLAQSHSLCIGPMPTSTFPPISSQL----	LVTHTHLPGTSQS	750
OY	149	SPPSPPHTDPELELVPRILSKHEELIQNMDRVDRITWVEQQISKUKKOQOOLEEAAKP		208
Dd	751	T--SGGITTPSSOPPLMSPVRHNSMSSTRPASII-----	QTLRHOSVMP	796
OY	209	EPEKVSPSPLESKRSLVOIIYDE-----	NKXEAHAHRILEGPOVELPLYNQ	259
Dd	797	-PDVISIPPPPPLPTH-----DBMAAPRGTPERSSETW-----	-VLNSP	835
OY	260	PSDTRQYHENIKINOAMRKULYFKRNHNARKQWKQRCORYDOJMEALEKKVERIENN		319
Dd	836	PFGT-----PIQNLLTMPVPPPH-----	LIAATSGTHSVSS	869
OY	320	----PRAAKSKYA---EYEKQPPERIKQRELOERNQR---	VGOGSGISMSA--AR	367
Dd	870	AHSTRHSISGVPACBSNSKTSPPTPKSR--BEKQIRIDTITSKGSPNAIALQAR		927
OY	368	SEHEVSEIIDGSEGENLEKQWRQOLAVIPMLYDADQORIKFINNN-----	GIMADPMKY	423
Dd	928	SQSMYS-----	GPCKSA PSTPYVDGDSLVQAQIMNQNLGIKRLPRIEK	974
OY	424	KDROYANMWSOEKETFEFEKFMQHBPKNFGLIASFLERTVAECVLYYLTKKNENYGLV		483
Dd	975	KSALONIQNHQPHSNASTPSTPSTSTHOAFMDXKE-----	RKKEKEKEER	1025
OY	484	RSRYARRKSOQOOQOOQOOQOOQOOQMPRSSOEKDEKEK--EAREKEEKPEVNDK		542
Dd	1026	EPEARREKREKETEERKKRKMEXAKGLEBRQRRKEKKEKERBERKEKEKYKKAEK		1089
OY	543	EDLKKEK-----TDDTSGENDENEKAVAASKRKTANSQGRKGRITRSMANEANSEAI		596
Dd	1086	EKLKKKHKRKGDSDSDSDSDDELIDLVRKS-----	TKEMQE-----	1122
OY	597	TPOOSABELASNELNSSRWTEEBMETACKGILLEHRMNMSALARNGSKTVSQCKNFYNY		656
Dd	1125	--EKDHQALL-----LSKGIGE-----	NL	1143
OY	657	KKRONLDEILLQOHKLKNEKENNARNARKKKAAPALAASEAAPPVVEDEMEASGVSGNEEE		716
Dd	1144	KSRRSRDGRANDSPBKQOKSQOKRVLIES-----	SDDGGXDGDGN--	1187
OY	717	MVEBAELHASGNEVPRGEGCGPATVNNSPTESI PSHTEAAKDTGONGPPRPATLGAD		776
Dd	1188	-----SNGEE-----SDSEK-----	AD	1199
OY	777	GPPGPPPTPPRTSNAPIETPPASEBATGAPPPRAPPSAPRPVPRKEEKEETAAPR		836
Dd	1200	LPPP-----	PAPPSISEADQRLKYLKERKEGELTTSSDD	1233
OY	837	VEEGEOQPPAABELAVDTGAEBPVNSECTEABEGBAGK-----	KDAEAEA-	885
Dd	1235	-EDHNDAGEIHQOQLTEDRENRRKKQSLTAYSBOGERKNUPPKMRHRDSEDAAAKHPG		1293
OY	886	-TAGGALKAEKKEGGSGATTAKS--SGAPQSDSATCSABEVDAEBCDNKRLSPRS		943
Dd	1294	WSAKDOQOKRKLEHRRSSDESCKNAKRFDI PHEDVSDEEBTEBGSSRSKROSTST		1355
OY	944	LITTEGPBRAVASFOKPLDKQLQORAAAIPPIOVTKNEBPREDDAAVTKAPRAPAPPQ		10013
Dd	1354	ISNVATAKERKKSXTPLRIIVPEBTGPRLSPKLSRPHLSPKTSTSTKXS-----		14005
OY	1004	NLOPESDAPQOGSSPRGKSR-----SPAPPADYEAPAAEQULPGDPBPCTSGLPFFVP		1058
Dd	1406	-----SISDHBNLISPQRNRNTSTSTATTSSGH-----BALSIDPEK-----	PLS	1446

453 HSTARRRGRSRRRRSRSDRRGSS-----HRSRREKRRR- 491
QY 1262 SLPKGHVIEGKKGHVLSYEGMSTQCKEDGRSSGPPHETAPKRTYMEGRVGR 1321
Db 492 -----RRSASPP-----A 500
QY 1322 ISSASIEGLMGRAPPERHSPHHLKEQHNRGSIQGIIPRSYVEAOEDYLREAKTLKRE 1381
Db 501 ASSSS-----SSRREH-----RGRREG-----GKKKKSRSAEGRS 535
QY 1382 G-----TPPPPPSRDLTEAYKTOALGPLKLPANBGLVATYKAGRSIHEIPRELRHTP 1437
Db 536 GDLEKLPAPVPPSGSDRSGRGAVP-----SIQDLTDHDLFAI- 575
QY 1438 ELPLAPRLKESITQGTPLKYDTGASTGSKKDVRSLISPGRTFPPVPL-DYMDA 1496
Db 576 -----KRTIVGRPDKTEPRAP-----SPA--PAVSPKGEVLVDS 608
QY 1497 RAL---ERACYEESLSKSPGTASSSGSIARGA--PVIPELGKPR-----OSPLTYEHNG 1547
Db 609 EGLADERGAOKDKORRRSGAASSSSSRERKSRKALDGRDRDRSSKPKTPKO-S 667
QY 1548 APPAGHLRGSFVTRKEEPTPLQESLSSSSKASQDKLTSTPRE---IAKSPHSTVPEH 1604
Db 668 APGSALPKAP-----RSGSSSSSSSSCSKXKLOSKVAVLIREGVSSTTPAKD 717
QY 1605 PHPLSPYHLRQVSGVDLYR--SHIP-LAFDPTSIPIRGIPLDAAAYLYLPHILAPPT 1660
Db 718 SSSSG-----LGSIGVFSRDRBSRSPFLKPDERSPAEGV-----KVAIPST 759
QY 1661 YPHLYPVLIGYPTALENNQTIIINDYITISQCHNHNTATMAQRAQMDLRGLSPRESSL 1720
Db 760 KPKTKVAKAKAGAKAKGTGK-----TKPSKTRKVRSGSGSTAGSGPSLKKSKAD 812
QY 1721 ALNTAAGRGIIIDLSQVPHLPVLPPTGTPATAMDRLAYLETPAPQSSSRHSSPLSPG 1780
Db 813 SCSQAASAKGTETSWSGE-----BRTKAPSTPPKVAAPPPALTPDSQVDSCKTP- 866
QY 1781 GPTHLTGTTTTSS-----SERERDRERDREREKSLITSTTVENAHPIWRPGTEOSS 1835
Db 867 DVSFLAERASBDTVRGVGAHEEEEEEEEEEOQPATTATSTAAAPSTAP----- 920
QY 1836 GSSSGSGGGGSSSRPASHNAHQSPISPRQDLOQRPVLAHTGKGIITAVEPSPK 1895
Db 921 -SAGSTAGDSGAEDGPAAVA-----SOLPT-----LPPMP 950
QY 1896 TYLRSTSSPVRPATPPATPCPLGTLGTVPTIMEPVLLPKFAPRVAPERRA- 1953
Db 951 NWL-----PAGVDCITTSGL-----ALTALLFKMEANLA--SKAKQE 987
QY 1954 ---DTGHAFL-AKPPARSGLEPASP-SKSEPRP--LVP--FVSGHATTIARTPAKNL 2002
Db 988 LIQDNTQILHRKPPSTLGVTPAPVPTSGLPAPSSYLLPGLSLIGCGGSPPTPT-GL 1046
QY 2003 APHAASPPRAPPA-----SDPHREKTSKRFPSIQELELRSLGN-GSYSPREGEVPS 2057
Db 1047 VP--ASDRKREGSSSEGRGDTKYLKTLTORAVEEVLSTIKPYQKKDITKEEYKIL 1104
QY 2058 PVSSTSLTHDKG---LPKHLLELDKSHLE-----GELR-PKQPGVKLG 2097
Db 1105 KRAYVKKICHSGKGEINPVKVSUVAIVYKRYFRKNGKPKDPGPPRPAPKEPBPDKG 1164
QY 2098 GEAAHLPHL 2106
Db 1165 GPGLPLPL 1173

RESULT 46
G01763
atrophin-1 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G01763

R:Margolis, R.L.
submitted to the EMBL Data Library, March 1995
A:Reference number: G08343
A:Accession: G01763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1184 <MAR>
A:Cross-references: EMBL:U23851, NID:g915325, PID:g915326
C:Genetics:
A:Gene: GDB:DRPLA, B37
A:Cross-references: GDB:270336, OMIM:125370
A:Map position: 12p-12p

Query Match 2.8%; Score 372.5; DB 2; Length 1184;
Best local similarity 21.3%; Pred. No. 4.2e-06;
Matches 331; Conservative 136; Mismatches 605; Indels 485; Gaps 75;

QY 762 TGONGPKPATLGADGPPGPPPTPRRTSRAPIDPTPASEATGAPTPPPAPSPSAPPV 821
Db 3 TRQKKDSMSMSGKKKAPGPRELRSGRASPGVGSTSSSDG----- 45
QY 822 VPKEKEETTAADPVEBGEQKPPAABELADTGAAEYPVKSCTEBAEGPAKQDAE 881
Db 46 --KAESRQTAKKARVEASTPK-----VNKQSRSEISESE--SEETNAPKKTTEE 94
QY 882 AAEATAGALKAKEKGGSGGANTAKSGAPQDSATGADAV-----DEAG 931
Db 95 LPPQSPEDL-----DSLGRSLNDGSDPRDIDQNRSTSPSISYFSGVENDSDSSG 149
QY 932 GDKRLSPRSLTPTGDPANASPOKPLDLKOLKQRAALPIQVTKVNEPREDAAP 991
Db 150 LSGPARRYHPPPLFPPEPQRPDSTPQP-----EASFPHSGVTPGYNAPEPPTSR 203
QY 992 TKPAPP-APPPOULOPE-----SDAPQPG-----SSPGRKSRPPAPDKAFA 1036
Db 204 MFOAPRGARPHNPQLYPGGTGVLSPGPMGKGGAASVCGPBGCKONPPPTPISVS 263
QY 1037 AEAQKLPDPPCWSGILFPVPRPREVIKAS--PHAPDSAFSYPARGHPLGLGHTARP 1094
Db 264 SGA-----SGAPPTPPTTPVGCGNLPSAPPPANPHNTPLP----- 301
QY 1095 VLPPTTISNPPLISSAKHPSVLERQIGALSQGVOLHVPYSEHAKAPVPTMGPL 1154
Db 302 -----PPPALRLPAMNASAPPLGA-----QPLGHLR-SPHA--MGQIGG--L 341
QY 1155 PMDPKK--LAFSGVKQEOQLSPRGQAPPSLGVPTAQEASVLRGTALSGVPGSITKG 1211
Db 342 PPGEPKPTLAP-----SPHSILPPASSAPAPPMRFYSSSSS--SSAASSSSSSSSSA 395
QY 1212 IPSTRVPSDSAITYRGSTTHGTADVLKGTITIIIGDSRSRLDRGKEDSLPKGHVYE 1271
Db 396 SP--FPASQALP--SYPHSFPFP-----TSLSVSNQPKY--TOPLPQAV-- 436
QY 1272 GKKGHVLSYEGMSTVQCKEDGRSSGPPHETAPKRTYMEGRVGRATSSASIE- 1328
Db 437 -----WSQGP-----PPPPY-----GRLLANSNAHGP 460
QY 1329 -----GLMGRALPPERHSPHHLKEQHNRGSIQGIIPRSYVEAOEDYLREAKTLKREGT 1383
Db 461 FPPSTGAQSTAHPPV--STHH--HH-----HQOQOQOQOQOQOQHNGN 500
QY 1384 PPPPPSRDLTEAYKTOALGPLKLPANBGLVATYKAGRSIHEIPRE-----ELRHTP 1437
Db 501 SGPPPP-----GAPPHPLEGSSSHNAHPVAMSSLSGLRLRYP 537
QY 1438 ELPL-APRLKESITQ---GTPLYKYDTGASTGSK-----KIDVRSL--IGSPGRTFP 1485
Db 538 PGPALHPPPHSGVSTISQAGPNGPPVSSSSSSSTISQGSYCSHSPGQGGAP-YPFP 596
QY 1486 PVHPLDVADARALERACYEESLSKSPGTASSSGSIARGAPVIVPELGKPROSFLTYED 1545
Db 597 PV-FVTTTSSA-----TLSTVIATVASSPAGYKTASPPGPPYGGKAPSGAAYKT 645

Db 785 EKGIAGKPGILPGLTGA-----KGEAGYPPGPG----- 811
Qy 1280 YEGGMSVTCCKEDGSSSGPPHETAPAKRTYDMWEGVRAISSASINGLWG-RAIPE 1338
Db 812 ----LPGKBPBGSTTGPPEPGPG- -LGKDG-IPGAGLPGLSQKPLP- 857
Qy 1339 RHPHLLKEQHHIRGSIITQGIIPRSYVEAOEDYLRRKAKLLKREGTPPPPSRDLTEAVK 1398
Db 858 -----GVPGQ-----KSGILPGLAGAPGF----- 878
Qy 1399 TQALPGLKAPRAHEGLVATVKAAGSIHEIPREELHTEPLAP-PLKSGSIITQGP- 1456
Db 879 ----GAKGEPGLPGLPG- -KSGPGPGPGGAPGPGQKGP-EGILPG 918
Qy 1457 -----LKYDTGASTTSGKKHIVRSLLGSPGRTFPPVHP-----LDVMAADRALR 1500
Db 919 LPGVSGMKDGTG-----LPGVGLAGPGGPGPGQKGP- - 963
Qy 1501 RACYESLSKRPQTASSSGSIARCAPVIVPELKGPRQSLTYEDHGAEPAGLIP- 1555
Db 964 ----EAGLPGLPGAPQKQKEGGLAGLPGIPKMGAPGIP-----GAGQDGLPGLPGV 1012
Qy 1556 ----RGSPTWREPTPLQEGSSSSKASQDRKLTSTPREIANSPHSTVEHHHPIS 1609
Db 1013 KCDRGFNGLPGEKGEFGPAPARDEKG-----EPGL 1042
Qy 1610 PYEHLLRGVSGVDLYRSHIPLAFDPTISIRGIPLDAAAYLPLRHLPAPTYPHLYPPLV 1669
Db 1043 PGQGRGPGQG-----PPGLP----- 1058
Qy 1670 IRGYPTALLENRQTIINDYITSQQMHNATATAMQORADMLKGLSPRESSLALNYAGPR 1729
Db 1059 ----GLGKMGDEG-----QPGYAGPLMGKKG--LGLGPKRPPRG--APGK 1099
Qy 1730 GIIDLSQVPHLVL-----VPTPTGTPATAMDRLAYLPAPQPFSSRHSSPLSPG--GP 1782
Db 1100 G---LDGAPGFPGLKGEAGLPAPGAP--GODGLPGLPG-----OKSGEFGPGQPLGVGP 1149
Qy 1783 THLTPTTSSSERDRDRERDRERKSLITSTTYEHAPIMRPGREGSSGSSG 1842
Db 1150 PGLP-----GKMGAPGIRGKGDGLPG 1172
Qy 1843 ---GGGSSSRPASHSHA-----HSHSPISPTODALQORPSVLNHTGMKGIITAVEPSK 1894
Db 1173 LFGEGGLDGLPGQKGEAGFPAPGAPGVPYKSGAGAPGF-----GLKG-----EPGL 1221
Qy 1895 PTVLSTSTSSVPRPATPPRATPCPLGSLDGVYPTLMEVLLPKREAPRVARPERP--- 1951
Db 1222 PG-LEGQPPRGMKGEAGLPGA--FGRDGLPGL--FGMKGEAGLP-----GLPGQPGKS 1270
Qy 1952 -RADTGHAFAPKPARSGL-----EP-----ASSPSKSEPRPLVPVSGHATTAR 1996
Db 1271 ITGPKNGMGLPGLPGKDGSLPGLPGLKGEFGKPGVAGAAIKGEPGLPGIPGAKGEPGSLG 1330
Qy 1997 TPAKNLAHPHNASPPDPAPPASADBPHEKTOGKPSIOLELPLSLGYHSSSYSPGVEVY 2056
Db 1331 IPGKR--GNDGIPKRPAGLPGP-----GMKGESEGLPQGGA 1368
Qy 2057 SPVSSPSLTTHDKGLPQHLELDKSHLEGLRKPQD-----PVKLGEBAHLPHL 2106
Db 1369 GLPGLPGLKGEAGLPGLPGQKGETGPGQ--PGIPGLPMKMGDSGVGAPRGDGAPEGKQ 1426
Qy 2107 RPLSSQSSSPPLQT-----APGVKHQVAVTLAQHSIVTIOTYTRHHHPQOLS 2156
Db 1427 BEGPMGPPEAOPITVQGEKGEWGPAGPGRGK----- 1460
Qy 2157 APLPAPLYSFPGASCPVLDIRRPSDLYLRPPDHGAPAR-----GSPHSGKRSDEPKNT 2212
Db 1461 -----GLPGLDGLPGPSG--PGFAKAGKGDGFPQGRPMGMEKGA- 1499
Qy 2213 SVLGG--GHDGIEPVSPPGKMPTEPGHSRSAAVPLLYRDG---EDTEPSRMGSKSPGNTS 2266
Db 1500 ----GLPFPPIEGIPGPPGLPGPSGPPGPGP--SYKQGFLLVKSQTSSEVYQCPFGMWK 1554

Qy 2267 QPPAFFSKLTESNSAMVSKQ-----EINKKLTNHRNEPEVNI 2306
Db 1555 LMDGYSLIYEGNE---KSHNODLGHAGSCLSRFTMTPLPCDVNNVNCVYASRNDKSYVL 1611
Qy 2307 SQPGTEIFNMPAITGTGLMTYRSQAVQEHASTNMGLEAIRKALMKYDQWEEPSPLSAN 2366
Db 1612 ST--TAPIPMFVSSGIEPIYSRCVACEAPAN-----VIAVHSQTIQIPN-CPN 1658
Qy 2367 AFNPLNASSLSLPAAPITRADORSHTLTPSGGGGKAKYSGRPSRRKAKSPAPGLASGR 2426
Db 1659 GNNSLMTIGYSF--AM-----HTGAGAGGGQS-----LSSPGSCLIEDFRA 1696
Qy 2427 PPSVSSVSEGEDCN 2440
Db 1697 TPFIENGARGTCH 1710

RESULT 45
T31421
C-terminal domain-binding protein ral - rat
C.Species: Rattus norvegicus (Norway rat)
C.Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
C.Accession: T31421
R.Yuryev, A.; Patrunajan, M.; Litington, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Cord
Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996
A.Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts wit
A.Reference number: Z21024; MUID:96293459; PMID:8692929
A.Accession: T31421
A>Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-1173 <YUR>
A.Cross-references: EMBL:U49056; NID:g1438531; PID:g1438532; PIDN:AA052657.1
A.Experimental source: hippocampus

Query Match 2.8%; Score 372.5; DB 2; Length 1173;
Best Local Similarity 21.5%; Pred. No. 4.2e-06;
Matches 311; Conservative 154; Mismatches 537; Indels 447; Gaps 69;

Qy 753 SPHTAAKDGTQNGKPPA-----TLGADGPPPPPPPPRTSPAPLEPTPASATGAP 806
Db 77 SPQSHASR-----PACSRHLTLTGDCGPAPPPAPSGSSSP-SPSPSSSPBP 126
Qy 807 TPPAPPSAPPP---VVPKEKEEETAAAPVEEGEQKPPABEILAVDTGKAEPPVK 863
Db 127 PPPPPPPALPAPFDIYDPHPTDEAYSPPAEQKYDPEFATGSPSSSGTSPSE 186
Qy 864 SECTEABEGPAKGDAEAMATAGALKAEKKEGSGRATTAKSGAPODSDSATCSA 923
Db 187 EEEEEE-----BEGLSQSIRRISETLAGIYDMSLSQDPFG 225
Qy 924 DEVDEABGDKNRLSPRSLTPTGDP-----RANASPOKPLDKOLKQRAAIIPIQ 977
Db 226 DDSPHRE-----PPPPQTGAPGTPOADSTABEAGPRRVAVVQPEAEACLEGKVS 277
Qy 978 VTKVHEPPREDAPTKPAAPAPPPONLOPESDAQO--OPSSRP-----GKSHSPAP 1029
Db 278 V---EVVTTAGGRLPLPPLP---TDPELEBSIYQPEEFPVAVSLRAAPRQPP 329
Qy 1030 ADKEAFAAEQCLPGDPCCWTSGLPFPVPREPVIKASPHADPSAFYAPGHPPLGLH 1089
Db 330 ASVATLASVA--APAAPA-----SAPRAPEGDF-----LSLH 361
Qy 1090 ----DTLRPV-LPPPTITSNPPPLISAKIPSVLEROIAGISQMSVQLHVPYSEHAKA 1143
Db 362 ADSDEGALQVLDGEPRA---PPAADARWGGLDIRRKI-----LTQREERYQORSGAS 410
Qy 1144 PVGPVTMLPLPMDOKLAPSGVQKQLSPRGAQGPESIGVPTAO--EASVLRGTALGS 1202
Db 411 P-GP-----PPARKKR--REKORS--GDPAPDS--PWEAKKGRSRERKIGS 452
Qy 1203 -VPGSITKGIPTSTVPDSAIYVGSITHGTPADVLYKGTITRIIGDSSPRLDGRGD 1261

```
Qy 1519 GGTATGAGPVIPELCKRQSLTYED-HGAPRAGHLPRG--SPWTMEPTPLQESLS 1575
Db 2015 SGSRRS-----RTSKRSQPL--EDLDGFQELFQTPAGSNPVSVESEA---KLSLE 2062
Qy 1576 SSKASQDKLTSTPRELAKSPHSTVPEHHPHPISEYHLLGVSGVDLYRSHIPLAPDPT 1635
Db 2063 SSQAEPTVTRASTKR-LSKT-----GLNMMDVREHSHPLS----- 2096
Qy 1636 SIPRGIPLDAAAYVLPRLHAPNPTYPHLYPPYLIRGYDPALENNQTIINDYITSQOM 1695
Db 2097 -----KSCASQGVMTLTIGEDH-----GRETKQKVLAAQXL 2130
Qy 1696 H---HNTATMAQRADMLRGSLSPRSSSLATVYAGPRGIIDLSQVPHLPVLPVPRG--T 1750
Db 2131 EPAIVYTRGKRQQRCKKRSQSPEDLS-----GVQEVFQ-----TSGHNK 2170
Qy 1751 PATAMDRLAIVPTAQPF-----SSR-----HSSSPLS----- 1778
Db 2171 DSVIVVDLAKLPSSSPLEPTDVTSTRKQARTGARKHVKNELSGIMHPQISGEIYDL 2230
Qy 1779 PGGPTHLYKPTTSSSERERDRDRERDREREKSIITSTTVEHAPIWR-----PGTEQ 1833
Db 2231 PREBEGEKVIKTRKQSVKRLDTEVNVPRSKRQRTIAEKTLEDLPGFOELQAPSLVM 2290
Qy 1834 SSGSSSGSGGGGSSSR---ASHSHAQ-----HSPISPR-----TODALQGRPS 1876
Db 2291 DSVIVETKPKMPDKSPEDVDTTSETQARRRLRLVLTVEEPIPOKRTTRVAVQTRNTQEP 2350
Qy 1877 VLHNTGMKGI-ITAVEPSKPTVLRS-----TSTSPVPAATPPATPCPLG--- 1923
Db 2351 ISDNQMEFKESSQKODPSVSLTGRNQPTVKEKQPLEELTSFOEETAKRISSSSP 2410
Qy 1924 -----TLDGYVPTLM-----EPVLLPKKAPVAVP--EPRADTGHAFIAPPA 1965
Db 2411 QPEKETIAGLKRQIQLINDGVKEEPTAQRKQSRRETRNTLKEPVDISINVEVKST 2470
Qy 1966 RSGLEPASPCKSGSPRLVPSVSHA-----TIART-----PAKUL 2002
Db 2471 KOKIDPVASVPSKRRP-RVPEKKAQALELAGKPIQTLGHTDSASDKGTQMPKNSL 2529
Qy 2003 APNHASPPRAPPASDPRHEKTO-----SKPESIQELERS-----LGYN 2044
Db 2530 QPEQVD-----SFQSSFRPRRTKRGVYEADEEBSAARKVYSTGRQTRSKKVEIGNN 2582
Qy 2045 GSSYSPEGV---EPVSPV--SSPSLTHDKGLPKHLEL--DKSHLEGELRPKOPGVKL 2096
Db 2583 GTQVSKASIKQTLDTVAKVTSGRQLRTHKMGSTLLKLGDSKEI-----TQI 2631
Qy 2097 GGEAHLPH---LRPLPESQPSSSPPLQIATAGVKGHQRVVTLAQHISEVITQDYTRHP 2152
Db 2632 SDHSEKLAHDTSTILKSTQOQKPSVYKPLRTC-----RRVLRKASEVEKEVLVD--TRGHA 2684
Qy 2153 ---QQLSAPLPAP--LYSPFASCPVLDLRPRPSDLYLPPPDHGAPARSPHSE---GGKR 2205
Db 2685 TLQGSNLLSPKRSADGSI VTRALRS-----LAPKQEAATBEKVPEKKRAASSKR 2738
Qy 2257 MGSKSPGNTSQPPAFSKLTES-----NSAMVSKSKQEI-----NK 2292
Db 2787 SRYRKTVVQPRKPFDAEENVGKKKQKTKTKTASQETELQNPDDGAKKSTSGQVSGK 2846
Qy 2293 KLNTNHN-----EPEYNISQPTETI FNNP 2317
Db 2847 RTCLRSRGTTEMPCPEAEKTSKRAAILIKP 2879
```

RESULT 44
S16366
collagen alpha 2(IV) chain precursor - pig roundworm

```
C:Species: Ascaris suum (pig roundworm)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: S16366
R:RefSeq: U.; Kingston, I.B.
J: Biol. Chem. 266, 16149-16156, 1991
A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the parti
A:Reference number: S16366; MUID:91340768; PMID:1714907
A:Accession: S16366
A:Molecule type: mRNA
A:Residues: 1-1763 <JBI>
A:Cross-references: GB:M67507; NID:G159648; PIDN:AAA18014.1; PID:G159649
C:Genetics:
A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfid
F:1-26/Domain: signal sequence #status predicted <SIG>
F:127-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F:127-42/Domain: non-collagenous NH1 #status predicted <NH1>
F:43-1529/Domain: collagenous #status predicted <COL>
F:197-199/Region: cell attachment (R-G-D) motif
F:1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1530-1638/Domain: repeat NC1 #status predicted <NC11>
F:1639-1763/Domain: repeat NC1 #status predicted <NC12>
F:31,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 2.84; Score 373; DB 2; Length 1763;
Beet Local Similarity 20.7%; Pred. No. 6.2e-06;
Matches 400; Conservative 119; Mismatches 673; Indels 742; Gaps 93;

Qy 727 SGNEVPGE-----CSGPATVNNSSDTEISPS-----PHTKADQNGQNPX 768
Db 299 SGAPGPKGSRPGLPGPKGDRGLDGLGVPLGQCKEAGFPGRDGA--GARGP 356
Qy 769 PPATLG--ADGPPGPPTPPRTSRAPLEPTPASEATGAPRP-----PAPSPSAPPV 821
Db 357 GPPGGERSDG--PPGPGPLGR-----BGQPPPADQVPGPPGPGQGL 401
Qy 822 V-----PKKEKEETAAAPVEEGEGRPPAABELAVDTSKAEPPVKE 865
Db 402 PGCGPLPGLPGLPGLPGKKGDSGIPGAPGVG---PPG---LGGPPAKGEP--- 450
Qy 866 CTBEABEGPAGKDAEAEAT--AEGALKAEKKGSGGATTAASSGAPDS----- 915
Db 451 -----GP-RGVGQSTPLPGKDGRLDGLPGRKKEGMLPGVRGPGDLSLGLRPP 502
Qy 916 -----DSSATCS--ADEVDAAEGDKN--RLISPR--- 941
Db 503 GPRPGQPKYGDGDGAPGLPGTIPGRKDRGTCTGAFCHAKGKGKGDAGVAGLPGQGER 562
Qy 942 --PGLTPTGPRANASPOKPLDKQLKQRAAIPRIQVTVNRPREDAAPTKRAPAP 999
Db 563 GLPGIPGATGAPGDDGLPGAP-----GRPPSP-----GPPGQDGLPGLPGQKE 606
Qy 1000 PPPQNLQPSAAPQPS-----SPRCKRSRAPAPADKFAFAEAKLPGDP-----P 1047
Db 607 FTQLTLRP--GPPYPGQKGTGPPRPGQGLPGKE---IVGAPGLGPPPPKGP 659
Qy 1048 CWTGGLPFP-----VPPREVIKASPHAPDPFAFSYAPRGP--LPLGLHDTARPLYPRPT 1101
Db 660 GLT-GLPEKPKQDGI PGLPGKGP-----GYGQGMGGLGMMKMDALPGLPGLG 710
Qy 1102 ISNP--PPLISSAHPVSVLEQIGALISQMSVOLHPVSEAKAPVGVMTGMLPMDPK 1159
Db 711 AVGPWGP-----VPPESQLRPP--PGKDGLPGLPBPX 741
Qy 1160 KLAPSGVQKQSLSPRQAGPPESIGVPTAQEASVLTALGSVPGSGITKGPSTRVPS 1219
Db 742 GEAGFPGAPGLQ--GPAGLPGLPKGNP-----GLPGAPGLAGLGG--IFG 784
Qy 1220 DSAITYRGSITHTPADVLYKGTITRIIGEDSPSLRDLGRGSDSLPKGVHIEGKKGHVLS 1279
```


A:Reference number: Z20787; MUID:96431717; PMID:8834799
 A:Accession: T30249
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2938 <STA>
 A:Cross-references: EMBL:X82786; NID:g1177527; PIDN:CAA58026.1; PID:g1177528
 A:Experimental source: strain CBA; testis
 C:Genetics:
 A:Gene: KI-67
 C:Keywords: cell cycle control; nucleus; tandem repeat

Query Match 2.8%; Score 373.5; DB 2; Length 2938;
 Best Local Similarity 19.4%; Pred. No. 16-05;
 Matches 576; Conservative 351; Mismatches 1073; Indels 973; Gaps 145;

QY 43 QHNSRDYASHLSPGSIIPORRRPSILSEFQGNER--SQBLHLRPSHSLYPLBLGKSE 99
 DB 182 QHTGNRIIVPTSGGSL-----RSPGLQAV--TGNRSILPTQSLSNSNEKESPFELKYGSM 236
 QY 100 MFFIESKR-----RLLELDPDL--RSPPLATG---QPAQSEDLTKDRS--LTKL-- 145
 DB 237 KEELDYKQSKCRKSPQPDRAAESRETQLLVSGARAKSSGSTFTVTAAASPVKGLWT 296
 QY 146 -----EPVSPSPPTDPELELVPRSLKEELIQNMDR-----VDREITM 185
 DB 297 EKMKGAMVVGQSTETAKKKTTPRHQOLKDEDSRVTGRHGVNLDGSSAQAVHKTVTP 356
 QY 186 VEQIISKLKKQOOLEE--EAAKPE--PEKVPSP-----PRIEKHR--SLVQIIT 231
 DB 357 -----GKLATRNQTPVEADVGSPADTPHSSSPQASIPAKVEAPFAELQNLSTLQRLV 411
 QY 232 DENRKAKEAHLRLBELGQVELPLYNOPS-----DTRQYHNNIKINQARKKLLILYFKR 287
 DB 412 PEKKTPKSPSKPEKLATPAEQTCGLFGLSSVDLSNFGDSTINSEGMPMK-----RKR 466
 QY 288 -----NHAKQWKQFCQRYDQMLEALEKVERIENPPRRAKESKVEYEKOFEPIRQ 343
 DB 467 VSRGHLRP-----ELFDENLP-----PMTPLKRGETPTKSKLGTSHSPAVLK- 509
 QY 344 RELQERMOS-----RVQORGSLSM----- 363
 DB 510 TIIKERPOSGKQESPGITPPRTNDQRRSGRTSSGNFLCETDIPKAKRGKGNLPAKR 569
 QY 364 -SAASEHEVSEI-----DGLSEQENL-----EKOMQOLAVIPMLYDAD 403
 DB 570 ASISRSQHGLQWICSKRRSGASEANLIYAKSWADVVKLGCVKQTOTKVAKHVPKQTSKR 629
 QY 404 QOR-----IKFINMGIMADPMKVYK----- 424
 DB 630 QRRPSTPKKPTSLNHNQFTTGHANSPCTIVVGRAQIEKVS-----PARPYMLNNLML 683
 QY 425 DRQV-----MNMWSEQEKETFEKFMQHPKNFGLI--ASFLEK-----TVACVLV 469
 DB 684 NRKVDSEDLGSLTEFMKTPVVEKEQOQMSDPTGVLNSANLSEKQLQVNSGDIPPIIT 743
 QY 470 YLITKKNENYKSLVRSYRRRGSKSQOQOQ-----QQOQOQOQOQOQPMRPSQSEKDEK 524
 DB 744 EILIGER-----VLSTRMAAQOSDRYSASPTLRRSLIKHENTVOTPKVHNITDLEK 796
 QY 525 KEKEAKEKEKEPEVENDKEDLLEKTDITSGEDNDEKAAVASKGRKTANSQGR--RKGRI 583
 DB 797 KIPVSTELPKLTSSVSKLRSRLNHTLVETMNEKTEAVLANTARHLRGTFRRQKVD 856
 QY 584 RSMANPANESEALTPQOASALAMEINNESSRWTEBEMETAKKGLLEHGRNWSAIAMVGS 643
 DB 857 QOVODNENA-----PQRCKE--SGELSEGEKTSARRSSARK-----QKPTKOLIGS 901
 QY 644 KTVSQCKNPFYFNKRR-----QNLDELQOHKLMKEMERARRKK----- 683
 DB 902 QWVOTQADVAEELLISQOQGITONLEBSEHMONTSISBDGITEKKNIIYATKEKSPK 961
 QY 684 ---KKA-----PAASEEAAP-----PVVED-----EEMASGVSG 712

DB 962 TPQKKAQPLEGPAGLKEHFETPNPKXETEDRRTRVLCKSPQVYNTENTITNTKQDQTSISG 1021
 QY 713 NEEVVEAEAL-----HASGN-----EYPRGCGSG-----PATVNSSDT--- 748
 DB 1022 KKVDMKEBSALYTRIKHMPGSRHNPKILKLECEDIKALKOSEMMLTSTYNGSKRTTGK 1081
 QY 749 -----ESIP--SPHTE-----AAKDTGONG-----P 767
 DB 1082 SKKKAQPLEDLTCQELFISVPVNTIIKITPSKSPHTQPVRTASTKRLSKTGLSKVDVR 1141
 QY 768 KPPATLG-----ADGPPGPPPTPP-----RTSRAPRIPT 797
 DB 1142 QEPSTLGRKTKSPGAPGTPPAVQENDCTAYMETPKQKLSIENLTGLRQOSRTPKQIT 1201
 QY 798 PASATGATPPRAP-----PSPSAPPVVPRKEKEEETAAAPVVEGEQKRP 846
 DB 1202 GFQDSFQIPDHANGVLVVKTKKQNFNSPQ--PESAITKSRBSGRAS-----ISKID 1253
 QY 847 AAEBLAVDTGKAEBFVKSECTEAEERPAKGDMAEATGALKAKEKGGSGRAVTA 906
 DB 1254 VKEELL-----ESEENLQGBGVDTFQVSTNKVIRSSRKPAKRLDSTA 1297
 QY 907 KSSGAPQ-----DSDSATGSAD-----EYDEABGGRKRLSPBSLL--TPGDP-- 951
 DB 1298 ---GMPNKRKMKSCSKDNTPCLEDLNGFQELFQMPGVYANDSLTGTISTMLRSPQLGVR 1354
 QY 952 -RANSPQKPLDKOL-----KORAAIPIQ-----VTKVHEPPRED---AA 990
 DB 1355 TQINKKSLPKITILRKQDVTEELSGLMKQSLGRVHTTQOEDNAIATIMEIKETIQTAD 1414
 QY 991 PTKPAPAPAPPPONLQPSD-----APOQSGSPKGRSPAP-----PA 1030
 DB 1415 GTRLRQQTPEKEXQPLBDSHFQELFQTSRYCSDPLIGNKQMTSLRSPQGFVTRPR 1474
 QY 1031 DKEAPA-----AAQKLP--GDPFCWTSG--LPPVPPRE-----VTKASPHAPDDSA 1074
 DB 1475 TSKRLAKTSVGNIAVREKISPVSLPQCATGEVYHPIPEBDTENKGVKESPTQTLDSA 1534
 QY 1075 -----FS-----YAPGHPRLPGLHDTARPVL--PRPTISNP 1105
 DB 1535 SKTVSKRQOGAHEERPOFSGDLFHPQELFQTPASGKDPYTVUDETNIKILQSPQHIIINP 1594
 QY 1106 P-----PLISSAKHPSVLERQIGALISQMSVOLHVPYSEHAPAVGVYTMGLPLPMD 1157
 DB 1595 ASMKRQSNMSLAKDMRPSILEKQ--TQSRGDA-----GTPAPMQ 1633
 QY 1158 -----PKLAPF-----SGVQOQLSRGQAGRPESLGVPTAOEASVLRGTAIGS 1202
 DB 1634 EENGTATIMETPKOKLDFIGNSTGHRRRPTPKNAQPLEDL--DGFQE--LFGTPAGA 1688
 QY 1203 VPGSITKGIPTSTRVPSDS-----AITVRSITHTPADVLKGTITRIIGEDSPSRID 1256
 DB 1689 SDPVSVESAKISLASSQAEPRTPASTKRRKKTGLSKVDV-----KQRPSTIG 1737
 QY 1257 RGRBESLPGKHVLYEGKKGHVLSYEGCMSVTQCSKED---GRSSGGRPHETAPRYTYDM 1313
 DB 1738 K-RKMSLGRA--PQTPAPVQENDSTAFMETPKOKLDFITNSGSHKRRPQTPKIRAQP 1792
 QY 1314 MEGRVG-----RAISSASIEGLMGRALIPREHSPHNHLEQHNGISITQGIPIRSV 1364
 DB 1793 LEDLDGFOELFQTPAGANDSVIVESVKKSLSSQAEF--VTPPASTKRLSTGL--SKV 1848
 QY 1365 EAOEDYLRREALKREGTPPPPSRDLTEAYKT--QALGPKLKAHIEGLVATVYKEAG 1422
 DB 1849 DVREDSIILEK--TSPGTPAVQEBNDCTAFMETPKOKLDTGNSGSHKRRPRTPK 1904
 QY 1423 RSIHEIPREELNHTBELPLAPRPLKESIT-----QGTPLKYDTGAST----- 1465
 DB 1905 ---IRAQPLEDLGQFQELFQTPAGASD--SVTVESAKMSLESQAKPVK--TPASTKRLSK 1959
 QY 1466 TSGKXHDV--SLIG-----SGRTTPPPNHPLDVADARALERACYEBSLKRPGTASS 1518
 DB 1960 TGLSKVDVREDSITGKTKTSPGRA--PQTPAPVQEBN--DSTAFMETPKOKLDFEANS 2014

A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1494 <CAO>
 A;Cross-references: EMBL:AF077000; NID:g3598973; PID:g3598974; PIDN:AAC62959.1
 A;Experimental source: brain
 C;Genetics:
 A;Gene: PTP-TD14
 C;Function:
 A;Description: may be involved in regulating Ha-ras-dependent cell growth
 C;Keywords: phosphoric monoester hydrolase

Query Match 2.8%; Score 373.5; DB 2; Length 1494;
 Best Local Similarity 20.3%; Pred. No. 5e-06;
 Matches 363; Conservative 200; Mismatches 619; Indels 609; Gaps 84;

```

Qy 51 SHLSPGSIIPQRRRLISLSEFQGNERSQF-LHLRPESHSLYLPDLGSEMEFTSKRPR 109
Db 72 AHLHMGKQAEQKQFGEVAYFQSLDLKLNBAIKLAKGQPDVQDALFADVIGGKXNS 131
Qy 110 LELLPPDLLRPS-PLATGQAGSEDLTKDRSLTGKLEPVPSPSPHTDPL-ELVPR 166
Db 132 AKKNDFTYHNAVPLDTLQPVKGAPLVPL-----PVNPTDPAVTGPDIFALVPE- 182
Qy 167 LSKELIQNMDRVNREITMVEQISKL--KKKQQLSEEAAPREPEKPVGPPPIESKGR 224
Db 183 -----MAAHNASLSYSEKAKILREMLAK-----IEDKNE 212
Qy 225 SLVQIYDENRKKAAHRIIEGLQPVLEPLVNPSPDTRQYHENIKINQMRKKLILYF 284
Db 213 VL-----DQFMSMQDLPDPTVDLDAV- 234
Qy 285 KRRNHAKQWKQKQFQRTDQLEALEKVEERLENNPRRAKESKRYEYEQPEIR- 341
Db 235 ---NHIPQMLEK-----CAALSVRPDVKLVQSMOVLGFTVDSLKQIRDL 283
Qy 342 KORELQF-RMGSRVQSGSGLSMSAARSEHVESEIIDLSEQENLEKQROLAVIPMLY 400
Db 284 EEDLQEOKLOETLQAGAGGEPSTYKAB-----LGEVREKHA-----KY 323
Qy 401 DADQQRIFINMGIMADPMKYKDRQVNMNMWSEQEKETFREKFMQHPKFGILASFLER 460
Db 324 TEVHEKASFNS-----ELHRAMNL-----HYGNILRLISGPIQ 357
Qy 461 KTVACVLYYLTKNGENYKSLVRSYRRKSGSQQQQQQ-----QQQQQQQQQPMRPSQ 517
Db 358 --VRAALPTPALTPED--KAVLQNLKRIILAKVQEMRQORVSLQEQLELLQKODITASL 412
Qy 518 EEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 576
Db 413 VTTTHSEMKKLFEEQK-----KYDQLKLYLEQNLAAQDN----- 447
Qy 577 RRRGRITSMANSEBAITPQOAGELASMEINSSRWTEHEMETAKKGLLEHGRNWSA 636
Db 448 ----VLAL-TEANVOYAAVRVLSL-----DQKWN-----ST 476
Qy 637 IARNVGS-----KTVSQCNFFYNYKKQGNLDILLQHKLAKHEKRNARK----- 682
Db 477 LQTLVASYEAYEDLMKKSQEGKDFYADLESK--VAALLERAQSLCRAGAAQQLDLDEL 534
Qy 683 KKKAPAAASEBAFPVVEDEMEASGVSGNEEM-----VEEAAALH 725
Db 535 KKKAP-PRPTPKPLLSRREGEAAAGADQPEELRSLPDMMAQPRLLPDPPLGTAAPLH 593
Qy 726 ASGNEVPREGSGSPATVNNSSDTESI PSHTEAAKDQNGKPRATLGADP----- 778
Db 594 FSPGFPDS--TGPAHYLS-----GFLPPGTV--SGPTOLMQR 629
Qy 779 -----PRGP-----PRRTSRAPITPASEAAGATPPAPSPAPAPV 821
Db 630 AAVPAAPGVLYPAVYTSGLVPRSSPQHGISPPA--GVGPPQPTVGLPSAPAPQ 686
Qy 822 VPKEKEEBETAAPVEEGEQPPAAEELAVDTKAEKPVSECTEAESEAPGAKDAE 881

```

```

Db 687 FSGPELAMD--VRPATTYVDSVQAPISSHMALRPGAPAD-POPCFVPQVVP----- 736
Qy 882 AAEATAGALKAEKEGSGRATTAKSSGAPQSDSSGATSGADVBDAEGDKURLSPR 941
Db 737 -----QSVPOPQPLPTYYTSI-----GTQKHLGPL 763
Qy 942 PSLTPTGDPANASPOKPLDLKQKQAAAIPIQVTKHEPREDAATKPAV-PAP 1000
Db 764 PQNHFPQITSPFAP-----RIGRPPRLQPPQPPQPPQPPQPPQPPQPPQ 812
Qy 1001 PPQNLQESDAPQPGSSPRGKSRSPAPADKFAABA----- 1039
Db 813 POPQPPQ-----PQRPVPGPQ-PTQQLPPLPQHPLPSPQAGGILPPPPPTPPQPVGLQ 868
Qy 1040 -----QKLPQDP-----PCWTSGLPPVPVPRVETKASPHAPDPSARVAPRGPPLGL 1088
Db 869 PPTPHQVLYGPPDPLPHSGALPPSP-----GPPH-PHPT-LAYGPAEPRPLGP 920
Qy 1089 HDTARPVLPPTTISNPPILISAKHPSVLERQIGALSQMSVOLHVPSEHAKAPVGPV 1148
Db 921 QAT-----PVISGRPPANQAPASPHL-----VPSAPSPGP-GPV 955
Qy 1149 TMSGLPMDPKCLAPFSGVQKQESLSPGQAGPPESLGVPTQOASVLRGTLASVPGGSI 1208
Db 956 PSRPPTAEPPLCLR-----RGAADAADLISSSPESQHG-----GTQPPGGG 995
Qy 1209 TKGIPTSRV-----PSSDAI-----TYRGSITHTPADVLTYKG 1241
Db 996 QPLQPTKVDAAERPTQALRLIEQDYEHEPERLOKLQELIESRFGQGDAGALDAVRE 1055
Qy 1242 TITRIIGDSSPSRLDRGDESLPKGHVYEEKGH-VLSYEGMSYQCSKED----- 1293
Db 1056 L-----QEAQENHARKSGTAAIAR--CYSLAKNHQDVMPTDSNRVLRSGKDYINASC 1106
Qy 1294 --GRSSGPPHETAPKR-----TYDMMEGRVGAISASITE-GLMGRALPPE 1338
Db 1107 VEGSLPCCP--LVAQRPPLGTADFWLWHEQKVGIVYVWLVEAEWEKQKVARYPEIE 1164
Qy 1339 RSHFHLKEQHNHIGSTIGCPRSYVNAQDEYLAREKL-----LKRE----- 1381
Db 1165 RGQP-----WVGAALS--VALSSVTRTDTHVRVLSLQRFQDSLKRSVLHLHPTEWE 1215
Qy 1382 -GTPPPPPSRDLTEAVYQALGPLKRAHEGLV-----ATVK--EA 1421
Db 1216 LGLDSP--GNLARFQEVNANHQRPLHTPIVNHCSGSGVGTGAFALLIYAIVQVEEA 1272
Qy 1422 GRSIHEIP-----REELRHTPE--LPLAPRLKSGSITQGPLKYDTGASTGSKKH 1471
Db 1273 GSRIPELPQVRRNRQOKHMLQEKHLKVLPRGI--GATCGAGPPAPARRAASQACGQH 1330
Qy 1472 DVRSLSIGRGTTPPVNPLDVMADARALERACYEBSLKSRRGTASSGSGSIARGAPVIVP 1531
Db 1331 E-ROLEPPSSGLGSGGW--WRRAHQHSQ-YHCQAIRP-----LGG----- 1370
Qy 1532 ELKPPROS-PLTYEDHGAFFAGHLPRGSPVTMRREPTPLQEGSISSKXASODRKLTSTPR 1590
Db 1371 -LDSBASLSVLVPRGIP-----PALPPTP-----APPSPP 1405
Qy 1591 EIAKSPHSTVEHNHPH-PISPYEHLKGVSGVDLYRSHIPLAPPTSIPIRG 1640
Db 1406 PSSLPPEPPQEEBPVSEAP--SLGPPSSLEILLASITPFAFSIDSLRG 1454

```

RESULT 43
 T30249
 cell proliferation antigen Ki-67 - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
 C;Accession: T30249
 R;Stardborg, M.; Geil, K.; Brundell, E.; Hoog, C.
 J. Cell Sci. 109, 143-153, 1996
 A;Title: The murine Ki-67 cell proliferation antigen accumulates in the nucleolar and he
 l for cell cycle progression.

atrophin-1 - human
 C:Species: Homo sapiens (man)
 C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
 C:Accession: S50832
 R:Nagafuchi, S.; Yanagisawa, H.; Ohaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.; Yam
 Nature Genet. 8, 177-181, 1994
 A:Title: Structure and expression of the gene responsible for the triplet repeat disorder
 A:Reference number: S50832; MUID:95144175; PMID:7842016
 A:Accession: S50832
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1184 <NAG>
 A:Cross-references: EMBL:D31840
 A:Gene: GDB:DRPLA; B37
 A:Cross-references: GDB:270336; OMIM:125370
 A:Map position: 12p13.31-12p13.3112p-12p

Query Match 2.8%; Score 373.5; DB 2; Length 1184;
 Best Local Similarity 20.0%; Pred. No. 3.9e-06;
 Matches 329; Conservative 128; Mismatches 588; Indels 603; Gaps 73;

520 KBEKEKEAEKEEKEPEVENDKEDLKKTDTGSDENDEKAVASGKRTANSQGRX 579
 2 KTRQNKDSMSGRKKEAPGREELRSGRASPGGVSTSSSDGAEKSRQTA-----KK 56
 580 GRITSMANNEASEAITPOQSAELASMLNESSRTEEMETAKKGLLEHGRNSALAR 639
 57 ARV-----EASITPKVKQGRSEISESE----- 80
 640 NVGSKTVSCKNFYFNKKRQNLDELQHKLMKERNARRKKKAPAAASEEAPPV 699
 81 -----SETVAPKKTK-----TEDELPRP 99
 700 VEDEMEASGVSGNEBEWEAEALHASGENVPRG-----ECSGPATVNNSSDTE 749
 100 QSPSLDLSL-----DGRSLNDGSSDPRIIDQDNRESTSIYSPGVENDSDSS 148
 750 SLPSPHTEAKDTGQNGPKRPATLGADGP--PGPPTPRTRSP--IEPTPASEATG 804
 149 SGLS-----QGRAP--YHPPPLFPSPQPDSTPRQPEASFEPHSPVTPTG 193
 805 --APTPP-----APSPSAP-----PVPVKEKEKEETAAAPVEEGE 841
 194 YHAPNEPPTSMFQAPGARPRHPQLYPGGTGVLSGPRMGKGGGAASVSGP---NGG 250
 842 ECKPRAAEELAVDTGKAEEPPVKSCTEEAEGBRANKDAEALAEALKAKEKGGSG 901
 251 KQHPPTPTTIVSSSGA----- 267
 902 RATTAKSSGAPQDSDSATCSADEVDEAGCDKNLLSPRSLLTPTGDPANNASPQKPL 961
 268 -----SGAPPTKPTTPVG-----GGN-----LPSAPPANFPHVTPMLPPR 305
 962 DLKOLKORAAALP--IQVTKNHP--PRADAAPTKRAPRAPPPONLQPSDADAOQGS 1017
 306 ALRPL--NNASAPPLGAQPLRGHLPSPYAMGQMGGLRPGREKPTLAP-----SHS 358
 1018 SPRGKSRSRDP--ADKFAFAEAOGLPDRPCWTSGLRFPVPRPVIKASPH--PD 1071
 359 LPPASSAPAPRMPRPVSSSSSSAAASSSSSS-----SSSSAPPPASQALPSYPHSFP 414
 1072 PSAFSYA--PGHPLPLGLHDTARVPLR-----PPLSNPP-----LSSAKGPSVL 1118
 415 PTLSSVSNQPKY-----TQSLPSQAVWSGQPP--PPYGLLLANSNAPPF 462
 1119 EROIGALISQMSVQLHVPYSEHAKAPVGVTVTGLPLPMDPKKLAFSGVKOEOLSPRQA 1178
 463 PPSGAGQTA-----HPVSTH-----HHHHQQQQQQQQQQQQQHNS 502
 1179 GPPELGVPTAOEAS-----VLKGTALGV--PGSITKGIPTSRVSDSALITRG 1227
 503 GPPPGAPRHPLEGSSHHNAPYAMSPSLGSLRYPVG-----PAHLPPHSGVSSQ 555

1228 SITGPRADVLKGTITRIGEDSPRLDRGEDSLPKGHV--YEGKKGHVLSYEGSMV 1286
 556 ACPNPGPV-----SSNSSSSTSGQ---SYPCSHPSGPGQCAPVPPPVPTV 602
 1287 TD-----CSKEDRSSGPPPHETAAPKRTYDMMEGRVGRALSSASIEGIMGRA 1334
 603 TTSSATLSTVIAVTAASSPAGVKTASPGRPPYCKR-----ASPAKYKA 647
 1335 TPERHSPHHLKEQHNGHISITGIPRSTVEAQEDYLREAKLKRBEGRP---PPPS 1390
 648 TTP-----GYKGGSPGF-----RTGPPRYGRTSPPA 675
 1391 RDLTEAVYKQALGPKLKP--HEGLVAVYKAGRSIHEIPRELHTEPLPLAPRLKGG 1449
 676 GGTGTRKPSPTVGPPLPPAGSGL-----PSLPPPAAPASGPPLSAT 719
 1450 SITGTPPKYDTGASTGSKKHVRSLLGSPGRT--FPVHPLDVNADARALERACYEEL 1508
 720 QIKQPAEYETPESV-----PPARSPSPPKVVDV--PSHASQSAFNNKL 765
 1509 KSRPGTASSSGSITARGAPVLYPELGK--PROSLTYEDHGAPFPAGHLPRGSPVTMRPT 1566
 766 -----DRGFNSCARSDLYFVPLEGSPWPRSGPTWRCGAGRAAR-- 808
 1567 PRLQSGSLSSSKASQDR---KLSTPREIAKSPHSTVEHHHPHPSPEHLLRGVSDV 1623
 809 ---KGAARARER 853
 1624 YRSHPLAFDPSTIPRGIP-----LDAAAYLLEPHLAP--NPTYPHLYP-----PY 1668
 854 ---RPPPEPSAAVATVPYLGPDTPALRTLSEYARHWSPGRNHPFVPLGAVDPG 908
 1669 LIRGY-----PDTALENQTINDYITSCQHNTATAMAQADMLRGISPRESSLA 1721
 909 LL-GYNVALYSSDPAARERERE-----ARERDRLRDKP----- 942
 1722 LNYAAGPRGIIDLSQVPHPLVLPPTGTPATMDRLAVLTPAPQFSRSSSLSPSG 1781
 943 -GFVKESEL-----EPLHGVPGGLD-----PF--PRIGGLALQGP 977
 1782 PTHLTKPTTSSERERDRERDRERERERERERERERERERERERERERERERER 1840
 978 PGLHFPFPHSLGPERER-----LALAAGPLRDMVAYEELAER 1019
 1841 -----SGGGGSSSR-----PASHAHQSPISFRTDALQORPSVLHNTGKIT 1887
 1020 QHAEVAGIGNDPLARLQMLNVTPHNHQSHISHILHODDAIHASASVHPL----- 1073
 1888 TAVERSKPTVLKSTSTSVRPAATFPRAATCPIGGLIDGVY--TLMEPVLPKEAPRV 1945
 1074 -----IDPLASGHLTIP-----YPACTLPN--LLP----- 1099
 1946 ARPERPADTGAFLAKPPASGLEPASPSPKSGEPRLVPPVSGHATARTPAKN--L 2002
 1100 -HPLHENVLHQFLAAYRDL---PASTLAPMSAHLQAMHMQSALQLALEQQWL 1155
 2003 APHNASPPRAPPASASDPHEKTOQKP 2030
 1156 HAHNPLHSVPLPAQEDDYSHLKESDKP 1183

RESULT 42
 T14355
 Protein-tyrosine-phosphatase (EC 3.1.3.48) TD14 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14355
 R:Cao, L.; Zhang, L.; Ruiz-Lozano, P.; Yang, Q.; Chien, K.R.; Graham, R.M.; Zhou, M.
 J. Biol. Chem. 273, 21077-21083, 1998
 A:Title: A novel putative protein-tyrosine phosphatase contains a BROL-1-like domain and s
 A:Reference number: Z18004; MUID:98361981; PMID:9649860
 A:Accession: T14355

Best Local Similarity 19.0%; Pred. No. 8.8e-06;
Matches 388; Conservative 249; Mismatches 799; Indels 607; Gaps 84;

```

Qy 15 TERYPRHSLSYPOIARTHTDVLGLVYOHHSRDYASHLSPGSIIO--PQRRPSLSSEF 72
Db 846 TMIYVPP-----PHLIAATSTGT-----HSVSSASHSTPRHSISGTPVACEPSNSKTS 893
Qy 73 QPCNERSGEHLRPS-----HSYLPGLGS-----EMEFIEBKRRLLELPPLLR 120
Db 894 QPPEPKS-----RPEKVOIRHDTTISKSGPSNAINALQARSQSMSTGDKKAPSTPVVD 948
Qy 121 --SPLLA---TGOP--AGSEDLTKORSLTGKLEPVSPSPPTDPELPELPRLSKEELIQ 174
Db 949 AGSDLVAGIMNQPRILGRLKPRIEKSSALQNIQNHPPHSNANSTSTSTSTHQMFM 1008
Qy 175 NMDVRDEITWEOQISLKKKKQOOLEEAAKPEPEKVPSPPIESKRSLLVQIYDEN 234
Db 1009 K-----DK-----EKERKKKEKEKEBERREARR-----EMRKETKE---ERN 1043
Qy 235 RKKAAMARILIEGLQPOVELPYNOPSPTRQYHENIKINQAMRKLLIYFRRNH----- 289
Db 1044 KRKEMERAKLEDERQERKREKREKDERKKEKVR--KKAKEKL-----KKKKRKGD 1098
Qy 290 -----ARQWOKQFCORYDOIMEALEKYEYRLENNPRRRAKESKYREYE 334
Db 1099 SDESDSDNDELDDYRKSTKEMTOEOKDQHLALLSKGIIENLKSRRSDKRAHDFE 1158
Qy 335 KQPEIRKQRELOEMQSGVQGR--SGLSMSAARSEHEVSEIIDGLSEQENLEKQWOLA 393
Db 1159 KMOQSGQRRVYIESSDEGQKDGKNSNGEESDEKADL-----PPP 1203
Qy 394 VIPPLVYADQORIKFIMNGLMADPMYVYKQRYOMNMMSDEKETREKFMQHPKNGL 453
Db 1204 PAPBLSASAQR-----LVYKEREKGLTSSDDHDNDAGEIHOQR--- 1247
Qy 454 IASFLERKTVACVLYVYLTKNENYKSLVRSYRRKSGSQOQOQOQOQOQOQOQOQOQ 513
Db 1248 -----LTEREN-----RKOKSLTAYSSDQGERKAVPKMR 1280
Qy 514 RSSQEE-----KOEKEK--EAEKEEKEPVENDKEDILKEKTDDTSGEDNDE 560
Db 1281 RDSDEDAAKHPGWSAKDDQOKRKRLKRRSSSEDESCKNAKDRPDIPEHEDVSDEFE 1340
Qy 561 KEAVASKRK-----TANSQGRKGRITRSMANSESEIITQ--QSAELASHEL 609
Db 1341 D---GSRRRQSTSTISNVTAKERKESGKTPRIPEPTGPLLSKRLISPKRLSEPT 1397
Qy 610 NESSWTBEEEMETAKKGLLEHGRNMSAIARWGSKTVSQCKNFYNYKKQONLDELQOH 669
Db 1398 STSS-----TKRSSISDH-----EULISPRQRN 1420
Qy 670 KLKKEKENARRKKKAPAAASEEAAPPV-----VEDE---EMASGVSG 712
Db 1421 R-TTSSSTATTSSGHEALSIPEKPLSPVTAKSSVSSIDPSIRDFSMNSAASPMT 1479
Qy 713 NEEVEVEEAELHAGNEVPR-----GEGSPATVNNSSD--TESI9PHTEAKD 761
Db 1480 TGRPVVLTKAAMKANSTPPKKKNSSGQHDSSSSSSSDSSDSTSDSDSDEVPKQ 1539
Qy 762 TGQNGPKRPATLGADGPPPPPP-----TPPRTSAPAPIERP--ASEATGATPPAPPS 814
Db 1540 T-----EYVTSIPVVASDNGSPENVVETPSIVSQTPREPEFTTISEGSSBEPFAVEBC 1594
Qy 815 PSAPPVVPKEKEKEETAAPRVEEGEQOKPRAABELAVDTGKAEVPYKSECTEBAEBCP 874
Db 1595 PEA--SVBPQMETSON--VEPVSEHDSHEHGDSEAVES--QOQPLEHQ--EKEKELE 1646
Qy 875 AKGQDAEA-----EATFEGALKAKKKGSGGRATITAKSSG--APQSDSACIS 922
Db 1647 NKILVAAEHHEEOYQGDSDSVSSIPAPSDPEPVTQAQEKSAITLLISDQTDQAVOSI 1706
Qy 923 ADEVDEAGGDKNRLLSPPSLLTPTG-----DPRANASPOKPL---DLKQLOKORA 970

```

```

Db 1707 FDEBEADE-----FPQYDPFGISTNEKEVSGKDPH--NIKTEPLNNGHTLLFSSSS 1758
Qy 971 AAIPIQVTKYHEPRRDAAPTRKAPAPPPQVLQOESDAPQOQSSPRKSKSPAPPA 1030
Db 1759 AAASEKQSTSEDEMEEDSELVNEKEV--PWEQVIAQEVHVSHP-----SPM-- 1805
Qy 1031 DKEAFAAEAGKLPDPCMTSGLPFPVPVPREVIASPHADPSAF-----SYAPPG 1081
Db 1806 -EEVVKLETSPVKEEPIKHEESPEQTPPLIINNNSQDTPGAVNNHLENHNAVQPI 1864
Qy 1082 HPLPLGLHTAPVLPPEPTT-----SNPPLISSAKHPSVLERQIGALISQ-- 1128
Db 1865 QLOPASHQOVAP--SPRPVAVPDSQNGPVLVSQSQPSPMWSQOSDMAQVLLISKDI 1923
Qy 1129 -MSVQLHVPYSBHAKAPVGPVTMGLPLPMDPKLAPSSGVYQEBL-----SPFGQAGPESL 1184
Db 1924 DLAALKHNPEALAOATRQDC-----SGIFQHLLLHAQNGQNMTPBML 1967
Qy 1185 GVPTAQASVYLRGTALDSVPQGSITKGIPTRPVSDSAITYRGSITHTPADVLYKGTIT 1244
Db 1968 QLKAAFPQOOQENANQOMQAKMQOITINDRIKEQERVAK-----MYEEN--E 2014
Qy 1245 RIIGEDSPSRLDGRBDS----- 1263
Db 2015 RKVEED--REKQKKEBERQLAAATAATATQKAAALQKQEVPRHGFQVLSMT 2071
Qy 1264 PKGHVITTEGKGHVLSIEGMSV-----TQCSKEDRSSGPHETAPART 1310
Db 2072 PEARSLYEOPFG--LSSYINRDSIGATNGVLHLPTQSIORPSTASTSNPKAPLOPAS 2130
Qy 1311 Y-----DWMEGRVGR--AISSASIEGLMGRAPPEBSHPHLKEQHIRGISTIOG 1358
Db 2131 VNQNTIPPAIEIEIRVQRMFVKPLKMSAEBAATMAVASSPNP-----PATSYVD 2181
Qy 1359 IPRSYVAQEDYLREAKLKRBEQTPPEPPSRDLTEAYKTQ-----ALQPLKL-- 1407
Db 2182 LAAMLQOLQAQAQAQVAVVVTASTPNLSLLETLLSTASLANLATGALNPLSMLA 2241
Qy 1408 -----KPAHEG-----LVATVGE-----AGRSIHETPR 1430
Db 2242 LTSSLNQSSPYQGIARVLLTMNQMOLATHQISELLATMNOBTLMALLAARNGLPFAM 2301
Qy 1431 EELHHTPELP-----LAPRLKEGSIQGTPLKYDTGASTGSKHDVRSILGSPGRTFP 1485
Db 2302 PQONOQPMQAGGFALPTVLPHMSLKRNADQLSVGVSBRKKSCLPHAMITGGQGOQPP 2361
Qy 1486 VPHPLDVMADARALERACYEBSLKSREPQASSSGSIARCAP----- 1527
Db 2362 PQQPMQAVAPA-----PPRSPSPPRKSMFENLPRMEKKNEMFRKELLRL 2407
Qy 1528 --VIVPELG-----KP--ROSPLYTD-----HGAPPAGHLPRGSPVTMRFP 1565
Db 2408 DIILIEELGAEDEBDQKDLQIPTSSEDTDSKADSGAGSAFRRLISS--TM--- 2462
Qy 1566 TPRLOEGSLSSKASQORKLSTPREIAKSPHSTVPEHHPRI--SPYEHL-----RG 1617
Db 2463 -----GNSSGPPASGTTSTSTSSISGCPDS--PRLEGPLBSFMDMLTEVAQKRE 2514
Qy 1618 VSGVDLYKSHIPLAFDPTSPRGIPLDAAAAYVLPRLHAPNTPYHLPVYLRIGYPTDA 1677
Db 2515 QSNTDALSAKI---VDEGSFQGHFPMWTGRLAKSTEA-----MINHLING----- 2559
Qy 1678 ALENRQITIINDYITSQOMHNTATAMORADMLRGLSPRESSLALNTAAGRGIIIDLSQV 1737
Db 2560 -----SETFLNDVLGRQVTEEN-----PRDSVXIL-----ORLRLDNGQV 2595
Qy 1738 PHL 1740
Db 2596 EHI 2598

```

RESULT 41
SS0832

QY 2295 NTHNNEPEYNISOPGTEIFNNPALITGTLMTYRSQ 2230
 Db 1534 NFASNDYSYWMISTEPMMMSQPLKGSIOFPISR 1569

RESULT 39

142731
 atrophin-1 related protein - rat.
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
 C:Accession: T42731
 R:Khan, F.A.; Margolis, R.L.; Loev, S.L.; Sharp, A.H.; Li, S.H.; Ross, C.A.
 submitted to the EMBL Data Library, December 1995
 A:Description: CDNA sequence and expression of an atrophin-1 (DRPLA disease gene) relat
 A:Reference number: 222250
 A:Accession: T42731
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1006 <KHA>
 A:Cross-references: EMBL:U44091; NID:g1297310; PID:g1209103; PIDN:AAA98970.1
 C:Genetics:
 A:Gene: ARK

Query Match 2.8%; Score 375; DB 2; Length 1006;
 Best Local Similarity 21.2%; Pred. No. 2.9e-06;
 Matches 268; Conservative 127; Mismatches 438; Indels 432; Gaps 59;

QY 546 LKEKTDNDSDGNDKEKVAASKRTANSQGRK-----GHITSMANEANSEALTPQ 599
 Db 5 VKEEDDGLSGKHS--MRTSRSGSMSTLRSGRKQSPADGR-----ASPVEDVRSGR 57
 QY 600 QSAELASMLNNESSRWTEEMETAKKGLLEHGRNMSAIAHWGSKTVSCKNFYFNKKR 659
 Db 58 NSPSASSTSSNDS-----KAAVYKS-----AKKYKEZAAPLKN-----TKR 95
 QY 660 QNLDEITLQOHKIKMEKENNARKKKKAPAAASEAPPVVEDEEMEAAGVSGNEEMVE 719
 Db 96 -----QREKVASDETDTRATSKK--TKTOEISRPSPERGESS----- 134
 QY 720 EAEALHAGNEVRECEGSPATVNNSSDTEISPH--TEAANDTON-----GPKPATLG 774
 Db 135 DSRVNDKSSDPK-----DIDQNRSTSPISPDNDSOSSSAQOQMLQTPPALQA 189
 QY 775 ADGPPGPTPTPRRTSRAPIEPTPASSEATGAPTTPPAPPSAPPVVPEKEKEETA 834
 Db 190 PGGAASASTAPPGT-----TQLEPTPTSATTVPPQSGP-----ATS 228
 QY 835 PVEBGEQKPPAAEELAVDTGKAEPEVKSCECTEAEBSGPAKQDAEAEAATAGALKAE 894
 Db 229 QPPNQTOSTVAPAAHTLLIQOTPTLHP----- 255
 QY 895 KKEGSGGATTAKSSGAGQDSSSATGADVEDAEGGDKNRLLSPRSLLTPGDPAN 954
 Db 256 -----RLSPHNPPLPMTAPPSQN 274
 QY 955 ASPQRP-----LDLKQKORAAAIPIQVTKVNEPREDAAATPKPAPRP 1000
 Db 275 SAQHPRQSLHGGQRPGRHSLQTGRLQHPGRPPQFGLT-----PQSSQCGSLGSPBALAH 331
 QY 1001 PPONLQ--PESAPQOQSSPRGKSRSPAPRADKEAFEAQOKLPDPRCWTSGLPFVVP 1059
 Db 332 PHSITQLPASQSALOPQOPPREQPLRPAPLA-----MHIKPPPTTPIF--QLPA 379
 QY 1060 REVTKASPHADPPAFS-----YAPRGHPLPLGHTARVVP-- 1097
 Db 380 PQAHHGPHALSGSPFSFNMNANLPPRPALKLSSSTHNPFAHPPPLQIMPOSGP--LPS 438
 QY 1098 --RPPTI-----SNPPLISSAKHPSVLERQIGALSQGSVOLAHVYSEHAKAPGVPTMG 1151
 Db 439 PAQPGGLGQSGLRP-----ASHPTT-----GGLHQ-----VPSQSPFPHPPVPGGP-- 484
 QY 1152 LPLPMDPKKLAPFGSGVKQEQSLSPRQAGRPESLGVPTAOEASVLRGTALGSGVPGSITKG 1211

Db 485 ---PIPPSCPTTS-----TP--PAGPSSSQPPCS--AAVSSG---GNVPGA----- 522
 QY 1212 IPSTVPS-----DSAI-----TYRGSTTHGTPTADVLKGTITTI 1247
 Db 523 -PSCPLPAVQIKKELIDAEBEESPDPSPSPSPPTVVDTPSHASQSAFXX----- 574
 QY 1248 GEDSPRLDGRDEDLPKGVHVEGKGKGVLSYEGGMSVTQCSKE---DGRSSSGPPHET 1304
 Db 575 -----HIDRG--YNSCATDLYFMPLASKLAKKKEEIEKAKRAEQAKAEERERREK 627
 QY 1305 AAPKTYDMMEGRVGRALISSAIEGLMGRAIIPPEHSPHMLKEQHINRISITQIPRGYV 1364
 Db 628 EKEREEREREEREAERAAQKASSAHGRLSDPQSGCHMRSPSE-----PPPTTI 679
 QY 1365 EAQGYILREAKLKRGETPPPPPSRDLTEAYKTQALGP-----LKLKPAHEGVA 1416
 Db 680 AAVPPYI-----GPDTPALRLSEVARPHVMSPTNRNHPFYMPINPT--DPLLA 726
 QY 1417 -----TVKEAGRSIHPIPREELRHTEPLAPRLKEGSIQTGTPLKVDYTGAST 1465
 Db 727 YHMGCLVAVDPTIRREIREREIREREIR--ERELRER--MKPG--FEVKPELDPLHPA 780
 QY 1466 TGSKKGDVRSLLIGSPGRTPPV--HPUDVADARALERACYEBSLKSPPGTASSGCSI 1522
 Db 781 TNPMEHFAR-----HSALTIPPAAGHPF-----ASPHGLNPLERERL 819
 QY 1523 ARGAPVIPELQKRRQSLTTEHGAPFAGLPRGSPVTMBEPTRLQ----- 1570
 Db 820 ALAGFQLRPE-----MSYPRDLAERIIAERMAASILT--SDPLARLQMFVNTPHNHQS 870
 QY 1571 -----EGSLSSSKASQDRKLSTPREIKSPHSYVPEHHPPI---SPYE 1612
 Db 871 HHSHTLHQQDPLHQSGAGVPHLV--PLTAGP--HLARFYF--FGTLPLPLGQPHNE 926
 QY 1613 H--LIRGVSVGLDYLRSHIPLAFDPTSIPRGIPLDAAAAYLP-----RHAPNPTY 1661
 Db 927 HEMLRHPVFGHPY-----PRDLPGALIPPMSSAHQQAQMAQSAELQRLAMEQOW 976
 QY 1662 PHLYP 1666
 Db 977 LHGHP 981

RESULT 40

T20532

hypothetical protein F07A11.6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T20532; T27777

R:Palmer, S.

submitted to the EMBL Data Library, October 1995

A:Reference number: Z19287

A:Accession: T20532

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2722 <MTL>

A:Cross-references: EMBL:Z65511; PIDN:CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b

A:Experimental source: clone F07A11

R:Gajdasty, S.

submitted to the EMBL Data Library, March 1996

A:Reference number: Z20417

A:Accession: T27777

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2722 <MT2>

A:Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b

A:Experimental source: clone ZK20

C:Genetics:

A:Gene: CESP:F07A11.6b

A:Map position: 2

A:Insertions: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3; 1

Query Match 2.8%; Score 374.5; DB 2; Length 2722;

A:Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 5
monomer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-
C:Function:
A:Description: minor structural component of extracellular basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAT1
F:27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status
F:27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F:42-1462/Region: interrupted helical
F:1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1473-1573/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1583-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
F:125/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
F:1527-1533,1638-1644/Disulfide bonds: #status predicted
F:1522-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

Query Match 2.9%; Score 377; DB 1; Length 1691;
Best Local Similarity 20.4%; Pred. No. 4.3e-06;
Matches 375; Conservative 136; Mismatches 647; Indels 678; Gaps 87;

Qy 764 QNGKPPATLADGPP-----PGRTPPRTSRATIEPPLPASEATGATPPAP-- 813
Db 143 QGPPGPPGIPGMKGGSIIMSSLPKGNPGYPPGPIQGLPGTGIIPGIPGPPGGL 202
Qy 814 -SPSAPPVV-----PKEKEETAAPAVEGE--EQKPAAE-----L 851
Db 203 MGPPEPPGPPKGMNFGQPKGKEQGLQGPBGPGQISEKRIIDVFQKDGGL 262
Qy 852 AVDTGKAEPYKSECTEBAEGRPAKDAEAAATAGALRAEKKGSGRATTAKSSGA 911
Db 263 PGDRGPPGPP-----GIRGPPPPGEGKEGKEQGEPPKRGKPDGNGQGI---PGL 314
Qy 912 PODSSATCADEVDVEEGDKN-----RLSPRP----- 942
Db 315 PGDPGPPGPPGPDG--EKQKQKDTGPPGPPGLVIRPRTGITTEKNGIQLPGLPEKGR 373
Qy 943 -----SLRPTGDP-----RANASPOKPLDK-----QLKORAAI 973
Db 374 GPPGIGGPPGPPGPPGAAMGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 433
Qy 974 P-PIQVTVKHEPPRED-AAPTKEPAPRPPPNLOPE-----SDAP 1012
Db 434 PGPBPAPRPHIPPSDEICEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 493
Qy 1013 QQPG---SSPRGKRSRAPPADK--EAPAAEQKLPGRPPCWTSLP--P--VPPRVI 1063
Db 494 GQPPGLPGLPBPGLSGFPQKQKQKQAGATGPKGLPIP--GAPAPGFPQSGKBPGL 551
Qy 1064 K-----ASPHAP-----DPSAFSY---APGHPPL 1086
Db 552 TTPPKGKGGKGLSGFAGLPGLPCTPGODGLPGLPGRKGEFGITTFGKGPFPANPGL 611
Qy 1087 GLHDAFVLPAPRPTISNPPLISSAKHPSVLERQIGAIISQMSVOLHVPSYEHAKAVG 1146
Db 612 GLPKNIGHM--GPPGFPGP-----VGEKGIQVVA--GNPQGPPIP-----GPKG 653
Qy 1147 -----PYMGLP-----LPMDPKTLAFSGVQKQQLSPROAGPPESLQVPT 1188
Db 654 DPGQITTPGKRGKGLGNRGDQVGLPDP-----GLPQPPGLPGLP 696
Qy 1189 AOEASVLRGTALGSPGSGITKGIPTSRVPSDSATITVRSITHTPADVLVYGTITRIIG 1248
Db 697 SKGEGTIGIGLPGPPG---PKGFPGITGPPGAP-----GTPGKIGLEG----- 737
Qy 1249 EDSPPRLDRGRDLSLPKHVITVEGKGHVLSYEGMSVYQSGKEDGSSSGPHETAAPK 1308
Db 738 -----PPGPPGPPGPKG-----EPGFALPGRPPGPPGLP-- 765

Qy 1309 RTYDMGKRGVGRATISSASISGLMGRAIPERNHPLKEQHNIRGSIQGIIPRSYVEAQ 1368
Db 766 -----GFKG--ALGPPGDKGFPGRPP----- 786
Qy 1369 DYLRREAKLKRSTPPPPPPSRDUTEAUYTOALGRPLK-----KPAHEGLVATYKEA 1421
Db 787 -----GRITLDLPGPKDVGPNQDGRPMGPPGLPGLVQVGGPPGPIGPIQGP 836
Qy 1422 GRSIHEIPRE-----LRHPELPLARPL--KESITQGTPLKTDGAST 1465
Db 837 G--LHGIPGEKGDPPGGLDVPPGPPGSGKIPGADPPIGPSS--PGLPGK--AGASG 890
Qy 1466 TSGKDHVRSILGSPGRTFPPVHPPLVDVADABALERAQYESTLSRPGTASSGSIARG 1525
Db 891 PPGTKGEM--GMMGPPG---PPGLGIPGNS--GVPLKDDGLQGLPGLPPTPEKSGSKG 944
Qy 1526 APVIVPEIGKPPQSLTYEDHGAFAGHLPRGSPVTVREPTPLQESLSSSKASODRKL 1585
Db 945 EP-----GLP-----GPPGPM-----DNLGSKKEKG-- 967
Qy 1586 TSTPREIAKSPHSTVPENHPHPIISPYHLRGVSGVDLYNSHTPLADPTISIPGIPIDA 1645
Db 968 -----EPGLPGLPVGSGPKGYQGLPDP-----GQP--GL 995
Qy 1646 AAAYLPLRLAP--NPTYPH-----LYPPYL-----IRGVPDTALENQTI 1685
Db 996 SGQGLPGRPPGKPNPGLPQGLIGPGLKGTIGDMGPPGQGVGPPGPPGPPGPPGPP 1053
Qy 1686 INDYTSQQMHNATMAQPADMLRGLSPRESSIALNYAAGPGLIDLSQVPHLPVL-- 1743
Db 1054 -----GSPGLPGQKD--KG--DPEISSIGLPLGPPGPPGPPGPPGPPGPPGPPG 1098
Qy 1744 -----VPTTPGP--ATAMDRLAYLPTAPQPFSSRHSSPLSPGPPHILTPPTSSSR 1796
Db 1099 SVCDPGLPGLPFTGAKGQPLPGFPPTGPPGPKGISGP--PQNPGLPGRP----- 1148
Qy 1797 ERDRDRDRDRREKESILSTTVEHAPI-----WRPTEQSSGSSGSSG--GGGSS 1848
Db 1149 -----GVGGGHPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1181
Qy 1849 SRPASHSHAHQHSISP--RTODALQORPSVLANHTGMKGIITAVEPSKPTVLRSTSSP 1906
Db 1182 GERQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1236
Qy 1907 VRPAATPPRATNCPGLGTLGVTTLMEVULLPKEARVVARPERPRADTGAFLAKPPAR 1966
Db 1237 PGPSPGPPALGPKGN-----PQPGP-----PGRP--GPTGQGLPGRPP 1277
Qy 1967 SGLPPASPSKSGSPRLVPPVSGHATITARTAKNLAPNHASPPRAPASDPHREXT 2026
Db 1278 PGL--PKNIGIKGKGNPQGLPGLPGL-----KDDQPPGLQGNPGR-- 1320
Qy 2027 QSKPPSIQELBELRSILGYHSSYSR--EGVEFVSFVSSPSLTHDKPLRNLBELDLSHLEG 2084
Db 1321 -----GLNMKGDPGLPQVPPFPKMKPS-----GVP-----GSAGBEG 1354
Qy 2085 E---LRKQPPRVKLGEBANLPHLRPLPSQSSSPLLOTAPVYKGNQVVTTLAQHISE 2141
Db 1355 EPGILGP--PGRPLPGLG---PQGSITIKGADGPPPIQOQPLKQ----- 1394
Qy 2142 VITQDYTHNHQQLSAPRLAPL--YSPPGASCPVLDLRRPSDLYLPPRHDGABARSGPH 2199
Db 1395 -----LPGQGGQGLPGLPGRPPD-----PGRKGLPFDAGKRGKDPG 1433
Qy 2200 SEGKRSPPNKTSVLG--GSEDGIEPVSPREGMTEPGHSRAVYVPLLYRDEQTEPSSMG 2258
Db 1434 LPG-----QPTRGIDGPPGPPGLQGPBGPPGTSSVANGF-----LTRHSGTLD--A 1479
Qy 2259 SKSPQNTSQPAPFSSKLTESNANVYKSKQ-----EINKKL 2294
Db 1480 PQCPQGLTQVVEGSSL-----YVQGNKRAHGQDLGTAGSCLRFFSTMPMFCNINNVC 1533

QY 2129 HQ-----RVTLAQHISEVITODYTRHHHPQULASAPLPAPLYSPGACSPVLDLRPP 2180
 Db 2612 EKGDVGFMPGRGLKGEERGVKACGIDGKGDKGEGRPP-----GRPLAQHKGEMGEPG 2665
 QY 2181 SDLYLPDPDHGAPAR-----GSPHSEGGKSPPEPNKTSVVG-----GEGG- 2221
 Db 2666 -----VPOGSGAPGKEGLIGPKGDRGPFQGPQPKDQGEKRGRTPIGLCGPFGSGNDGS 2720
 QY 2222 IEPVSP-----PEGMT-----EPGHSRNVAPLLYRDGEQTEPSRMGSSPGNTS 2266
 Db 2721 AGPPPPGSPVGRGREGIQQGQKGRGPPGERVYVGAPEVGAERGEQGRPCPAGPRGEX 2780
 QY 2267 QPPAPFSKLTESNSAMVSKKQKQKINAKLNT-----NRHEPYNISQPTETEFN 2315
 Db 2781 GRNA-----LTEDDI-----RGFVRQEMSOHCACQGOFLASGRPLPSYADVTASQULHA 2830
 QY 2316 MPAI-----TGTLMTRSQVQVCEHASTNMGLAIIIRKALMGKYDOWEESPP 2362
 Db 2831 VAVLVKSHAEERERPPEDDESEYSESVERYQDP-----EA-----PWDEDDP 2875
 QY 2363 ISANAFNPPLNAGASLPAPAMP--ITADGRSD--HTLTSPGGGKAKVSGRPSRRAKSP 2417
 Db 2876 GCL-----PLDEGSCRAYTLRWYHRAVTGSTTEACHPFVYGGCGGANNRFGTRACERRCP 2930

RESULT 38
 S22917
 collagen alpha 5(IV) chain precursor, renal splice form - human
 N.Alternate names: procollagen alpha 5(IV) chain
 N.Containers: collagen alpha 5(IV) chain precursor, leukocyte splice form
 C.Specter: Homo sapiens (man)
 C.Date: 30-Sep-1993 #sequence_revision 27-Feb-1997 #text_change 21-Jul-2000
 C.Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35
 R.Zhou, J.; Hertz, J.M.; Leinonen, A.; Trygvason, K.
 J. Biol. Chem. 267, 12475-12481, 1992
 A.Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identifi
 n Alport syndrome patient.
 A.Reference number: S22917; MUID:92316923; PMID:1352287
 A.Accession: S22917
 A.Molecule type: mRNA
 A.Residues: 1-967 <ZHO>
 A.Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234
 R.Zhou, J.; Leinonen, A.; Trygvason, K.
 J. Biol. Chem. 269, 6608-6614, 1994
 A.Title: Structure of the human type IV collagen COL4A5 gene.
 A.Reference number: A54365; MUID:94165049; PMID:8120014
 A.Accession: A54365
 A.Molecule type: DNA
 A.Residues: 1-922 <ZH2>
 A.Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AA027816.1; PID
 R.Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paape, A.; Trygvas
 Science 261, 1167-1169, 1993
 A.Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited sm
 A.Reference number: A57079; MUID:93361972; PMID:8356449
 A.Accession: A57079
 A.Molecule type: DNA
 A.Residues: 1-27 <ZH4>
 A.Cross-references: GB:Z37153; NID:G587203; PIDN:CA085512.1; PID:G587204
 R.Pihlajaniemi, T.; Pihlajaniemi, E.R.; Myers, J.C.
 J. Biol. Chem. 265, 13758-13766, 1990
 A.Title: Complete primary structure of the triple-helical region and the carboxyl-termin
 A.Reference number: A37122; MUID:90337990; PMID:2380186
 A.Accession: A37122
 A.Molecule type: mRNA
 A.Residues: 84-439; 'GS', 442-624, 'IALQ', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>
 A.Cross-references: GB:U05558; EMBL:M58526; NID:G1314209
 A.Note: submitted to the EMBL Data Library, February 1991
 A.Note: the authors translated the codon GCC for residue 115 as Val
 R.Renieri, A.; Serl, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma
 Hum. Mol. Genet. 1, 127-129, 1992
 A>Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in
 A.Reference number: I54317; MUID:93244772; PMID:1363780
 A.Accession: I54317

A>Status: Preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: mRNA
 A.Residues: 313-324, 'E', 326-330 <REN>
 A.Cross-references: GB:S59334; NID:G299946; PIDN:AAD13909.1; PID:G4261609
 R.Hoslika, S.L.; Eddy, R.L.; Byers, M.G.; Hoeykva, M.; Shows, T.B.; Trygvason, K.
 Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
 A>Title: Identification of a distinct type IV collagen alpha chain with restricted kidne
 A.Reference number: A34850; MUID:90160375; PMID:1669491
 A.Accession: A34850
 A.Molecule type: mRNA
 A.Residues: 914-1264, 1271-1691 <HOS>
 A.Cross-references: EMBL:M3115; NID:G180824; PIDN:AAA52045.1; PID:G180825
 R.Zhou, J.; Hoslika, S.L.; Chow, L.T.; Trygvason, K.
 Genomics 9, 1-9, 1991
 A>Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that
 A.Reference number: A37969; MUID:91169491; PMID:2004755
 A.Accession: A37969
 A.Molecule type: DNA
 A.Residues: 924-1264, 1271-1691 <ZH3>
 A.Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMB
 8; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:G177922; PIDN:AAA51558.1; PID
 R.Guo, C.; Van Damme, B.; Van Damme-Lombaert, R.; Van den Berghe, H.; Caessman, J.J.; M
 Kidney Int. 44, 1316-1321, 1993
 A>Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex
 A.Reference number: I56971; MUID:94133540; PMID:8301933
 A.Accession: I56971
 A>Status: translated from GB/EMBL/DBJ
 A.Molecule type: mRNA
 A.Residues: 1258-1276 <GUO1>
 A.Cross-references: GB:S69168; NID:G545095; PIDN:AA060612.1; PID:G545096
 A.Note: kidney splice form
 A.Accession: I76598
 A>Status: translated from GB/EMBL/DBJ
 A.Molecule type: mRNA
 A.Residues: 1284-1291, 'TLGLYACIV', <GUO2>
 A.Cross-references: GB:S69169; NID:G545097; PIDN:AA060613.1; PID:G545098
 A.Note: frameshift mutation in patient with Alport syndrome
 R.Myers, J.C.; Jones, T.A.; Pihlajaniemi, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; So
 Am. J. Hum. Genet. 46, 1024-1033, 1990
 A>Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the regi
 A.Reference number: A35335; MUID:90252791; PMID:2339699
 A.Accession: A35335
 A.Status: nucleic acid sequence not shown
 A.Molecule type: mRNA
 A.Residues: 1448-1477 <MYE>
 A.Cross-references: H.; Hattori, S.; Ushijima, T.; Matsura, T.; Koitabashi, Y.; Takada, T.; Yos
 Kidney Int. 46, 1307-1314, 1994
 A>Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord
 A.Reference number: I56975; MUID:95156893; PMID:7853788
 A.Accession: I56975
 A>Status: translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1595-1602 <NAK>
 A.Cross-references: GB:S75903; NID:G913882; PIDN:AA033374.1; PID:G913883
 A.Note: premature termination mutation from a patient with Alport syndrome; one other mu
 R.Lemink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Trygvason, K.;
 Genomics 17, 485-489, 1993
 A>Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo
 A.Reference number: I54188; MUID:94010948; PMID:8406498
 A.Accession: I54188
 A.Status: translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1604-1607, 'VHDAYKC', <LEW>
 A.Cross-references: GB:S65767; NID:G428563; PIDN:AA013967.1; PID:G4281667
 A>Note: frameshift mutation from a patient with Alport syndrome; five other mutations ar
 C.Comment: Prolines and lysines at the third position of the tripeptide repeating unit (e
 ed and subsequently O-glycosylated.
 C.Genetics:
 A.Gene: GDB:COL4A5; ATS
 A.Cross-references: GDB:120596; OMIM:303630
 A.Map position: Xq22-Xq22
 A.Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/
 /3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1165/1; 1

A>Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis
A>Note: there are 118 introns
C:Complex: type VII collagen is probably a homotrimer
C:Function:
A:Description: structural component of extracellular polymer associated with anchoring
C:Keywords: collod coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F:36-201/Domain: von Willebrand factor type A repeat homology <WMA1>
F:231-318/Domain: fibronectin type III repeat homology <FN1>
F:327-413/Domain: fibronectin type III repeat homology <FN2>
F:414-502/Domain: fibronectin type III repeat homology <FN3>
F:508-553/Domain: fibronectin type III repeat homology <FN4>
F:598-683/Domain: fibronectin type III repeat homology <FN5>
F:686-771/Domain: fibronectin type III repeat homology <FN6>
F:776-862/Domain: fibronectin type III repeat homology <FN7>
F:864-952/Domain: fibronectin type III repeat homology <FN8>
F:954-1045/Domain: fibronectin type III repeat homology <FN9>
F:1052-1219/Domain: von Willebrand factor type A repeat homology <WMA2>
F:1170-1172/Region: cell attachment (R-G-D) motif
F:1189-1253/Region: cysteine/proline-rich
F:1334-1336/Region: interrupted helical
F:1334-1336/Region: cell attachment (R-G-D) motif
F:2008-2010/Region: cell attachment (R-G-D) motif
F:2553-2555/Region: cell attachment (R-G-D) motif
F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F:2876-2929/Domain: animal knitz-type proteinase inhibitor homology <BPI>
F:337-786,1109/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:2167-2176,2185,2189,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F:2625,2631/Binding site: carboxydrate (Lys) (covalent) #status experimental
F:2634,2802,2804/Disulfide bonds: interchain #status predicted

Query Match 2.9%; Score 383; DB 2; Length 2944;

Best Local Similarity 22.1%; Pred. No. 4,8e-06;

Matches 437; Conservative 143; Mismatches 761; Indels 638; Gaps 110;

Qy 726 ASGNEVPRGECSPAT--VNSSDTEIPSPHTEAKDTGONGPK-----PPATLGADGP 778
Db 1303 ADGRGSPGRAGNPTGPAPGLKSGPLGPRGDPG-ERGRPGRGEFGAPQVIGGEGP 1361
Qy 779 -----PPGPTP-----PR-----RISRAPIETPASEATG 804
Db 1362 GLPGRKDDPPGSPGPRGLDPPGRPPGLPGTAMKGDCKDRERPPGEGEGIAPG 1421
Qy 805 APTPPAPSPSAPRPVVP--KEEKEETAAAP-----PYBEGRE--OKPPAAELAVD 854
Db 1422 EPLGLPGLSPGPPGPPGPPGPKKGSBDGAPGLPQPGSPGEGRGRPPGALIGPKD 1481
Qy 855 TGKAEPPVKSCTEBAEBGP-----AKGDAEAAATAGALKAEKKGSGGATTAKS 909
Db 1482 RG-FPGPL-GEAGEGERGPPGPASRGLPVAGRPAGK--PGPPPGPTGRGEGKGP 1536
Qy 910 GAPDSDSSATCSADEVDEAGCGDKRLISPP-----PELLPTGDPANANAP- 957
Db 1537 GRP--GDPAVVGPAVAGKGEKGDVGP--AGPRGATGVGGERGPPGLVLP--GDPGRKDPG 1592
Qy 958 -OKPLDQLQKRAAIPRIQTVKHEPPREDAATTKAPAPRPPPNLQPSDAAPQDG 1016
Db 1593 DRGFLGL-----TGAAGPPGD-----SGPPGKDGGRGPPGP 1626
Qy 1017 SSPRKSRSAPRPADKEAFAEAQKLPGDPCMTSGLPFPVPPREVITKASPHAPPSAFS 1076
Db 1627 VGRPRDDEVEKDEBP-----PGDP-----GLPGKAGERG-LGAPAVTGPVGRK 1672
Qy 1077 -----YAPGHPLPLGLHTTANPVLPPTTISNPPLISSAKHPSVLEQIGAI 1125
Db 1673 GDGDPGEGDGRNGSGSSGPKG--DRGEGPPGP--GRLVDTGPGAR 1716
Qy 1126 SQGMSVOLHVPPYSEHAKAPVGPV--TMGLPLPMDPKKALPFGSVKQEQSLP--RGOAGPRE 1182
Db 1717 EKG-----EPGDRQGBRGRKGPGLPGAPGERGIGBGRGPGRGQDPGVGRGAGEKG 1770

Qy 1183 SLGVPTAQEASVU---RGTLGSPVGSITK-----GIPSTR-----VPDSAAITYRGS 1228
Db 1771 DRGPGGLDGRSGGLCKRGKGAAPSGPNCAGAKGADPPGDDGLPGLRGEGGLPSPG----- 1824
Qy 1229 ITHGTADVLYKGTITRIIGEDSPSRDLRGREDELPRGHVLYEKGKHHVLSYEGMSVTQ 1288
Db 1825 -PPLPDKP-----GEDKRGKLNKNNGEDPGDGE--DGRKG-----EKGDG--G 1863
Qy 1289 CSKEDGRSSGPPPHETAPKRTTYDMEGRYGRAISSASISLMKRALPPRHSHHLKEQ 1348
Db 1864 ASGREGR--DGPKGERGAP-----GILDPGPPGILGP----- 1894
Qy 1349 HHTIGSTTGGIPRNVYAQEDVILREAKLKRSGTP-----PPPPSNDL--TEAY 1397
Db 1895 ---VGPPQGGPPVPGGTGPKDGEFGSKQEQGLPGERGLRGSPGVNVDLLLETAIGI 1951
Qy 1398 KTOALGPLKLPKPAHEGLVATVKEAGRSIHETPRELHTPELPLAPR--PLKEGSIOTGT 1455
Db 1952 KASAL-----REIVETDESSGFLVVP--ERRGRPKDSDGEGGRPKGEP--GF 1998
Qy 1456 P-----LKYDTG-----ASTGSKKHVRS-LIGSPGRTPPVHLVDVADALALERA 1502
Db 1999 PGERGLKGRDPPGPPGGLALGERGPPGSGLAGSPGKPGILPLP-----GRAAGVGEA 2054
Qy 1503 CYEBSLSRPRCTASSG-----GSITARGAPVIVELKPKRPSPLTYEDHGAAPFAGHLPRG 1557
Db 2055 -----GRPERGEREKGERGQGDGPPGLP--GTP-----GPPG--PPG 2091
Qy 1558 SPVTMRREPTPL--OEG--SLSSKASQDRKLTSTPEIAKPSHSTVENHP-----H 1606
Db 2092 PKVSDVEGPPGLSGEQRPGLKAKGKBPSSNGDQPKDQKRVPRGDKDRSGPRGQDGN 2151
Qy 1607 PISPYEHLRGVSGVDLYRSHIPLAFDPTSIPRGIPLDAAAYVLPRLHAPNPT-YPHLY 1665
Db 2152 PGLRGE--RGMAG-----PEGKPOLQPRGPPGPPGVS-----HGDPGPGAPGLA 2194
Qy 1666 PPLYLRGVPDIALENKOTIINDYITSGQMHNTATMAQAQDMLRGIPRESSIALNYA 1725
Db 2195 GP--AGQGPSGLKGE--GEGPPRGKGLTGTGAGLGLP 2231
Qy 1726 AGPRGIIIDSGVPHLPVLPPT--PGTP-----ATAMRLAYLPTAPOPFSSRHSSPLSP 1779
Db 2232 PPGSGLVPPQGSPLPGVGTGKRGAPRGAGAGKGDORSPPVP-----GSPGLP 2283
Qy 1780 GGPHTLTPRTTISSEERDRDRDREREKSI-LTSTTVENAP-----IWRPG 1830
Db 2284 GPVGRKBPFTGAP-----GQAVVGLPRAKGEKAPGGLAGDLVGBPG 2327
Qy 1831 TEQSSGSSGSG--GGGSSSRPASHSNAHQSPISPTQDALQORPSVLHNTGMKGIIT 1888
Db 2328 AKGDRGLRPPRGKEAGEARAEPPDPEGQAGAP-GPR-----GPKG-- 2368
Qy 1889 AVESKPTVLTSTSSPVRPAATPRATTHCPLGTLIDGVVPTIMEPVLLPKEARPVARP 1948
Db 2369 ----DDPGV-----GVPGSGPPGPVGVKDDL--LPG-----LGAGEVVGFP 2406
Qy 1949 ER--FRAD-----TGHAFIAKPPASGLERBASPSKSEBRPLVPPVSGHATTARTAK 2000
Db 2407 GQTPGRGEMGPPGSGERGLAGPPRGEGT-----PGILGP-----GPPG 2446
Qy 2001 NLAPHHAS-----PDP--PAPPASADPHREKTSKPPSIOELRLSIGYHSSISFP 2050
Db 2447 SVGPFGASGLKDKDPPGVGLPGRGERGEFGIINGEDRP-----GQEG--P 2491
Qy 2051 EGV--EPVS-----PVSSPSLTHDK-----GLPKHLELDKSHLEGELR 2087
Db 2492 RGLTGPPPSRGERGKGDVGSAGLKGKGDGSAVILGPPGPPGAKGDMERBRRLGIDGKG 2551
Qy 2088 PK---QPPVYKAGE-----AAHLPHLRPL--PESQPS-----SPLLQTAAPGVKG 2128
Db 2552 PRGDNQDPPGDKSGKEPDXKSAGLPGLRGLLGPQGGGAGAGIPLDPPSPGKDVGPGRIG 2611


```

Db 680 SPLHSGRFDMWKKSPILLRABRFTP-----SEASHRIPESVTLPSNRTSAGTSSSGVNR 734
Qy 106 KRPLELLEPDLRPLRPLATGQAGSBDLTKDSLTGKLEPVPSPPHDPELELVP 165
Db 735 KRKKRVSP---IRSEP---RSPSHMRTSGRLSSSELSPLTPSSVSSLSISVP- 786
Qy 166 RLSKEELIQNDVDRREITWVEOISKLKKOQOLEEBAKRPPEPEKVPSPPIESKRS 225
Db 787 -LATSALNPTTFPSHSILT-----OSGEAEKKQRPKKTSAPEAPFSSSSTPL- 835
Qy 226 LVQIITYDENRKKAAHRIEGLAPOVELPLYNOPSDTROGHENIKINOAMRKLLIFYK 285
Db 836 -----FPMWTFPSQTERGRNMDKAPSELSD- 861
Qy 286 RRRNARAKWKQKFCORDQMLEALEKVERLENNRRRAKESKYREYEKOPPEIRKORE 345
Db 862 -RDADKSVEXDK-----SREDREREKENRESREKER-----KKGSEIQSSEA 904
Qy 346 L-----OERMQRVQORSGLSMSAAR-SEHEVSEIIDG-----LS 380
Db 905 LYPYGRVSKKEVVEDATSSAKKATGRKSSSHSDSTDTITVTLGDTTAVTKILIKK 964
Qy 381 EOENLEKQMRQLAVIDPMLYDADQORIKFINNGIMADPMKYKDYKQYMMMSBOEKETP 440
Db 965 GRGNLEKTNLDLGPAPSL-----EKEKTLCLSTPSSSTVK-HSTSSIGMLAQADKLP 1018
Qy 441 REKMOMPKRNGLIASFLEKTVACVLYYLTUKNEYKSLVRRSYRRRQKSOQOQOQ 500
Db 1019 TDK-----RVASLTKKKAQOLC-----KIEKSKSL 1043
Qy 501 QOQOQOQOQOQMPRSSOEBKEKEK-----EKEAKEEBEKEVE- 539
Db 1044 -----KOTQOPAKQOQESSESTSVTCPRIKAVCPRAVALGAKRAVFPDDMFTLSAL 1096
Qy 540 ---NDKEDLKEKTDTSGEDNDEKAVASKGRKTANSQGRKGRITRSMANEASEAI- 596
Db 1097 PWEREKILT-----SSMGNDKSSIA--GSEDAEPLAPRIKPIKPYTRAKAPDEPVPK 1147
Qy 597 ---TPQOSAEIASHELNESSHWTEBEMETAKKGLIEHRMWSAIAKMGSKTVSOCKNFY 653
Db 1148 KGRSRRCGCGCGCOVPEDCVCTNCLDKPRFG--GRN-----IKKQC- 1188
Qy 654 FNYKRRONLDEILQOHLKMEKENARKKKAPAAASEEAAPRVVEDEMEASGVSGN 713
Db 1189 CKMKCCQMLQ--MMPSKAYLOKAKAVKKKKSKTS-----EKDOKSKSSVVK 1236
Qy 714 EEEVVEAEALHASGNEVPRGECG-----PATVNNSDTESIPSPH-----TEAAKD 761
Db 1237 ---VDSQKPTPAREDBPAKSSSEBPPRKPYBEKSEBGNVSAFGESKOATTPASRK 1293
Qy 762 TGQNGKPRPATLADGPPPPPT--PPRTSRAPLEPTPASEATGAPPPAPSPS- 816
Db 1294 SSKOVOSPALVI-----PPQPPITGPPRK--EVP-KTTPSEPKKQPPPEPSGEOQOK 1345
Qy 817 --APPVVPKKEKEEETAAPVEGE-----EOKPRA----- 847
Db 1346 KVARPSPIVAKQPEKEKEPPPVNKQENAGTLNLTSLTNGNNSKQKIPAGGVHRIKVD 1405
Qy 848 ---AEB-----LAVDTGAE-----EPVSECTEEAAE 872
Db 1406 KEDEAEVEMMGGLITSVITPRVVCFLCASSGHEFYCYCQCCPFFKFCLE-NE 1464
Qy 873 GPAP-----GKDAEAAEATAE-----GALKKEKE- 897
Db 1465 RPLDEDOENMCRRCKCFCHVGRQHOATKOLLECNKCNYSYHPECLGNYPTKPKTKKV 1524
Qy 898 -----GSGGATTAKSSGAPQDSDPSA-----TSADAEVDEAB 930
Db 1525 WICTKVCARCKSCGSTTPGKMDAOWSHDFSLCHDCAKLPAKGNFCPLDCKYDDDDYISK 1584
Qy 931 -----GQDKNRLISPRPSILTPT-----GDBRANASPOKPLDL 963

```

```

Db 1585 MMOCCKDURVWHSKCNILSDMEWELSNLPESVAVTCVNCTERRHPAEWRLALEKELOISL 1644
Qy 964 KO-----LKORAAALPIQVTKVHEPREDEAPATKAPAPPPONLOPES 1009
Db 1645 KOVITALLINERTTSHLRKRAKRPDL-----NPETEBESLPSHSSBEGDPVLTEYSK 1699
Qy 1010 DAPQPGSSPRGKRSRPAADKEAFAAEQKLPDPPCWTSGLPFPVPREVITKASPHA 1069
Db 1700 QKKOP-----LDLEGKRMQDQNTYSVLEFSDDIYKIIIOALINS 1740
Qy 1070 PPSAFSYAPRPHPLPLGLHDTARP-----VLP-----RPTISNPPPLISS 1111
Db 1741 DG-----GQBEIKKANSVMKSPFIQMERVFPWFVSYKRSFWEBENKSSNSMGJPN 1791
Qy 1112 AKHSVL-----ERQIGA-ISQMSVOLHAYSHHAKRPVCPVTMGJPLMPDPKLA 1162
Db 1792 AVLPSSLDHNYAQOEERENSHTEOPLMKKILP---APKPKGGEPSSTPLPPTPP 1847
Qy 1163 PFSGVKQEBLSPRQAGPPESL-----GVPTQABA----- 1192
Db 1648 ILSTDREBDSR--ELNPPGIEDNRQCALCTYGDSDSANDAGRLYIGQEWTHVNCAL 1905
Qy 1193 -----SVLRGTALG--SVPGSITKGIPESTRVPSDAITYRGSIT 1230
Db 1906 WSAEVPEDDDSLKNVMAVAIRGKQRCFCQKCGATV--GCCITSTSTN----- 1953
Qy 1231 HGTPADVLYKKTIR-----IIGEDSPSRIDGREDSLPKGHVI-----YEG 1272
Db 1954 -----YHFGCRANKCFPLDCKVYCOHRD--LIKGEVPENGEVEVRRVFDDEG 2003
Qy 1273 -----KKGHVLSYEGMSVT-----QCSKEDGRSSSGPP 1301
Db 2004 ISLRKFLNGLEPENNIMHIGSMITIDCLGILNDLSDCEDKLPITGYQS----- 2052
Qy 1302 HETAPKRTYDMEGVRGAISASIEGLMGRAIPERHSPHILKEOHNIRGSTTQGIPT 1361
Db 2053 -----RVWSTDAKRCVYTCKIIVECRPVVEPDINSTVEHDEKNTIAHS-----PT 2100
Qy 1362 SYVAQEDYLAREKALKREBTPPPPPSRDLTAY-----KTOALGPLKMPAH 1411
Db 2101 SFTSSSKESQNTYELIS-PPSPDRPPHSQSGCYVHVISKVRIRIPSPSPORSQGC 2159
Qy 1412 EGLVATYEAGRSHEITPREELRHTPELPAFLRLEKESITQGTPLKTYDTASTTGSKCH 1471
Db 2160 RPL-----PSAG-----SPTTHHEIYVGPDL-LSSGLRISGRRH 2195
Qy 1472 DVRSILGSPGRTFPPVPHLDVMADARALERAQYEE-SIKSRPGTASSSGSIAAGAPYIV 1530
Db 2196 STSSL--SPQRS-----KLIRIMSPMRTGNTYSRANVSVSSTGTAT-----DLESSAKYVD 2244
Qy 1531 PELGKPROSPLTTEYDHGAPRA-----GHLPKRGSPVTMRREPTPLQSGSLSS 1577
Db 2245 HVLDG-PLNLSSTLSQNTSTSSNLTORTVTVGNKNKSHLDGSSSSBKOSASADLVSKSSL 2303
Qy 1578 KASODRKLTSTPRRIASPHSTVPEHHPHRIPSPREHLIRGVSVDLYRSHIPLAFDPTSI 1637
Db 2304 KGEKTKVLSKSSG--GSANHVAYRGIP-KIAPQVH--NTTSRELNVSKISGSAEPBSV 2357
Qy 1638 PRGIPLDAAAAYYLPRHLAPNPTYPHLYPYLINGY-----PD----- 1675
Db 2358 -----SFSKSEAL-----SEPHIL-----LGGQRNDRDQHTDSTQASNSPDEDETEV 2399
Qy 1676 ---TALERNQTIINDYITSQ-----MHNTATNAAQRAMLRG----- 1712
Db 2400 KTLKLSGMSNNSIINHMGSSSRDROKGGKSKETPKERHSSKSLFLEQOVTTGEBGN 2459
Qy 1713 LSPR--ESSILANT--AAGPRGIIIDLSQVPHLVPVLPPTGPTATAMRLAYLPAPO- 1766
Db 2460 LKPEFMEBVLTPREMGQRPCNNVSSDKIGDKGLSMPCVPKAPRPMQOVGSSAELQAPRKRT 2519
Qy 1767 -----PPSR--HSSSPLS-----PG-GPTHLYTPTTSS 1793
Db 2520 VKVTLTPLKMEBESQNALKESVTLRPFANRVNISTEPISSAENPDGDPVAPQSPNNTSC 2579

```

QY	1164	FSGVQEOQLSRGQAGPESIGVPTAOASVYLRTGALGSVGGITTKIPISTRNVPDSOI	1223
Dp	528	HSGSEESDSEEEAAPAAS-----AAQAK-----PALEKOKAKSRKTPA-----	569
QY	1224	TYRGSITGTPADVLYKGTITIRIIGEDSPSRILDRGREDLPRKHVITYEKKGHVLSYEGG	1283
Dp	570	---SATGASTSSHCKAGAVTSSALSISP-----ALAKG-----	599
QY	1284	MSVYQCSNEDRRSSSGPHETAPKRTYDMMEGRGRAISSASIEGLMGRAIPEBRHSPH	1343
Dp	600	---TORSVD--SSSESESEGAAP--STRVQSGSG-----KGLQKALQGVAVP	645
QY	1344	HLKCOHNRIGSITQCIPIRSYEAQED--YLREAKTLKREGPP-----PP	1387
Dp	646	HTQTKG-----PSYKAMQEDSSLSLEDSSSEEDBETPAQATLGLPLQAKANP	695
QY	1388	PPSSDLTEAYVTOALGPLKILPRHEGLVAYVKEAGRSIHIEIPRELRHTPELPLAPRLK	1447
Dp	696	PTKTPPASASGKAVAPATPKGRPVVN--STVSARQO-----RSVPAAGKAGAPAT	743
QY	1448	EGSITQIGELPKYDTCGASTGSKNDVRLISPGRTFPVPRVLVMDARLYEACVYES	1507
Dp	744	Q--AQKPVV--GTGEDSSSSKESDEEBETPAQ---IKPVGKTSGVRAASAPAKESP	795
QY	1508	LK-SRPGASSGSGSIAGCAVIVPELCKPQSPPLYE---DHGAPAGHLP-----	1555
Dp	796	KKGNHFGIPKGTGSSATQAP-----QKTEDSBSSSESDTPEMBAQIKSPVSVN	849
QY	1556	RGSPLYMEKPTPLQEG--SLSSSKASODRKLSTPREIAKSPHSTVPEHHPHRISPEYH	1613
Dp	850	RNSPPAVAPAPPP--EGQAVNTTKKASGTTAOGSSSESDQEDLIPATOP--STVAL	903
QY	1614	LLRGVSGVDLVRSHIPLAFDPTS-----IPRGIPLDAAVYLPRLHAPPTYPHLYRP	1667
Dp	904	RTSVTTPALSRRA---ASQPSKESQSRMKPKAKAAMASAOQTSSAVETLPMMP----	954
QY	1668	YLIRGVPDTALENROTIINDVITYTSOQHNTATAMQORAMRLGLSPRESSILANTYAG	1727
Dp	955	-----PQSATIQPKAT-----NKLGSKULPEKQOALPGVPYKA	986
QY	1728	PRGIIIDUSQVHLFVLVPEPTGTPATAMDRLAYLPAPOPSSRHSSSPLSPGSPHTLTK	1787
Dp	987	PRSESDSDT-----SSEDEEDAKRPOWPK--SAHRLLP--DPSQETVVE	1028
QY	1788	PTTSSSRERDRDRERDREREKESILITSTTYEHAPVWPGRGOSGSGSGGGGGG	1847
Dp	1029	ETPRESS-----DEMVAPOSILISGYT-----PGLTVANSQASK-----A	1065
QY	1848	SSRPASHGHAQHSPISBPTODALQORPS-----VLHNTGMKGIITAVEBSKPTVLNST	1901
Dp	1066	TPRDSNSILA--SAPATKUNPDGKOKSKSQAUAUTALPKTGRKASSGSTQKPKLKK	1124
QY	1902	STSPVPVPAATFP-----PATCPLGDTGIDGYITLMEPVLLPKEARVARPE	1949
Dp	1125	TSSSPA--PTQILPNSITQRLLEQAMPSEAOVQSVYKVLTELLE-----QE	1170
QY	1950	RPRADTGHAFLAKPPASSGLEPASPSPGSPREPLVPVSGHATITARTPAKNLAPHHASP	2009
Dp	1171	RLKA-----TEATRESGKKSQKRLSGDLEGA-----PKN---KKK	1205
QY	2010	DPPAPASADPHREKTQSKPFSIQLELRLSGYHG--SSYSPGEGVPS-----EVS	2061
Dp	1206	EQVPBRASAVSPKAPMTSKAKS--KLDKSGAGKGKSPQPGAKEXKPDGELLGITLES	1263
QY	2062	PSLTHDGLGRPHLELDKSHLEGRLRQRPQPVVLGGEAHLPHLRPLRESQPSBP	2118
Dp	1264	GEQSDPFSKSKKKKSLKKK--DKERKEKKKKSLANDQAS--PLQKKKKKKKSAEP	1318

RESULT 36
A44265
trithorax homolog HTX, version 2 - human

N:Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-2003
A:Accession: A44265; A44264; A44263; A44262; A44261; A44260; A44259; A44258; A44257; A44256; A44255; A44254; A44253; A44252; A44251; A44250; A44249; A44248; A44247; A44246; A44245; A44244; A44243; A44242; A44241; A44240; A44239; A44238; A44237; A44236; A44235; A44234; A44233; A44232; A44231; A44230; A44229; A44228; A44227; A44226; A44225; A44224; A44223; A44222; A44221; A44220; A44219; A44218; A44217; A44216; A44215; A44214; A44213; A44212; A44211; A44210; A44209; A44208; A44207; A44206; A44205; A44204; A44203; A44202; A44201; A44200; A44199; A44198; A44197; A44196; A44195; A44194; A44193; A44192; A44191; A44190; A44189; A44188; A44187; A44186; A44185; A44184; A44183; A44182; A44181; A44180; A44179; A44178; A44177; A44176; A44175; A44174; A44173; A44172; A44171; A44170; A44169; A44168; A44167; A44166; A44165; A44164; A44163; A44162; A44161; A44160; A44159; A44158; A44157; A44156; A44155; A44154; A44153; A44152; A44151; A44150; A44149; A44148; A44147; A44146; A44145; A44144; A44143; A44142; A44141; A44140; A44139; A44138; A44137; A44136; A44135; A44134; A44133; A44132; A44131; A44130; A44129; A44128; A44127; A44126; A44125; A44124; A44123; A44122; A44121; A44120; A44119; A44118; A44117; A44116; A44115; A44114; A44113; A44112; A44111; A44110; A44109; A44108; A44107; A44106; A44105; A44104; A44103; A44102; A44101; A44100; A44099; A44098; A44097; A44096; A44095; A44094; A44093; A44092; A44091; A44090; A44089; A44088; A44087; A44086; A44085; A44084; A44083; A44082; A44081; A44080; A44079; A44078; A44077; A44076; A44075; A44074; A44073; A44072; A44071; A44070; A44069; A44068; A44067; A44066; A44065; A44064; A44063; A44062; A44061; A44060; A44059; A44058; A44057; A44056; A44055; A44054; A44053; A44052; A44051; A44050; A44049; A44048; A44047; A44046; A44045; A44044; A44043; A44042; A44041; A44040; A44039; A44038; A44037; A44036; A44035; A44034; A44033; A44032; A44031; A44030; A44029; A44028; A44027; A44026; A44025; A44024; A44023; A44022; A44021; A44020; A44019; A44018; A44017; A44016; A44015; A44014; A44013; A44012; A44011; A44010; A44009; A44008; A44007; A44006; A44005; A44004; A44003; A44002; A44001; A43999; A43998; A43997; A43996; A43995; A43994; A43993; A43992; A43991; A43990; A43989; A43988; A43987; A43986; A43985; A43984; A43983; A43982; A43981; A43980; A43979; A43978; A43977; A43976; A43975; A43974; A43973; A43972; A43971; A43970; A43969; A43968; A43967; A43966; A43965; A43964; A43963; A43962; A43961; A43960; A43959; A43958; A43957; A43956; A43955; A43954; A43953; A43952; A43951; A43950; A43949; A43948; A43947; A43946; A43945; A43944; A43943; A43942; A43941; A43940; A43939; A43938; A43937; A43936; A43935; A43934; A43933; A43932; A43931; A43930; A43929; A43928; A43927; A43926; A43925; A43924; A43923; A43922; A43921; A43920; A43919; A43918; A43917; A43916; A43915; A43914; A43913; A43912; A43911; A43910; A43909; A43908; A43907; A43906; A43905; A43904; A43903; A43902; A43901; A43900; A43899; A43898; A43897; A43896; A43895; A43894; A43893; A43892; A43891; A43890; A43889; A43888; A43887; A43886; A43885; A43884; A43883; A43882; A43881; A43880; A43879; A43878; A43877; A43876; A43875; A43874; A43873; A43872; A43871; A43870; A43869; A43868; A43867; A43866; A43865; A43864; A43863; A43862; A43861; A43860; A43859; A43858; A43857; A43856; A43855; A43854; A43853; A43852; A43851; A43850; A43849; A43848; A43847; A43846; A43845; A43844; A43843; A43842; A43841; A43840; A43839; A43838; A43837; A43836; A43835; A43834; A43833; A43832; A43831; A43830; A43829; A43828; A43827; A43826; A43825; A43824; A43823; A43822; A43821; A43820; A43819; A43818; A43817; A43816; A43815; A43814; A43813; A43812; A43811; A43810; A43809; A43808; A43807; A43806; A43805; A43804; A43803; A43802; A43801; A43800; A43799; A43798; A43797; A43796; A43795; A43794; A43793; A43792; A43791; A43790; A43789; A43788; A43787; A43786; A43785; A43784; A43783; A43782; A43781; A43780; A43779; A43778; A43777; A43776; A43775; A43774; A43773; A43772; A43771; A43770; A43769; A43768; A43767; A43766; A43765; A43764; A43763; A43762; A43761; A43760; A43759; A43758; A43757; A43756; A43755; A43754; A43753; A43752; A43751; A43750; A43749; A43748; A43747; A43746; A43745; A43744; A43743; A43742; A43741; A43740; A43739; A43738; A43737; A43736; A43735; A43734; A43733; A43732; A43731; A43730; A43729; A43728; A43727; A43726; A43725; A43724; A43723; A43722; A43721; A43720; A43719; A43718; A43717; A43716; A43715; A43714; A43713; A43712; A43711; A43710; A43709; A43708; A43707; A43706; A43705; A43704; A43703; A43702; A43701; A43700; A43699; A43698; A43697; A43696; A43695; A43694; A43693; A43692; A43691;

Query Match	2.9%;	Score 383.5;	DB 2;	Length 3968;
Best Local Similarity	17.9%;	Pred. No. 6,4e-06;		
Matches 565;	Conservative 357;	Mismatches 1088;	Indels 1247;	Gaps 143
OY	51	SHUSPSSTIQPRRRPPLLS--EPQGNRSGSEHLRRSHSYLP----	ELGKSEMFIFS	105

```
QY 1727 GPGGIDLS---QVPHLVLPVPTPGTATAMRLAY----- 1760
D 2003 SPSSFDASCKDSQSTAIIISPSPDRPHSQTSSCYTHYISKVPRIRTPSYSTQSPG 2062
QY 1761 ---LPTAPOF-----SSRHSPLSP-GGPTHLTPTTSS 1794
D 2063 CRPLPSAGSPPTTHEIVTGDPLSSGLKSGRRHSTSLSPRLKRLMSPVRIGSA 2122
QY 1795 EREDRDRERDREREKSLITSTTTEHAPIMRPTGSSGSSGGGSSSRASH 1854
D 2123 ---YSRHSVSVSLCTATDPEASAKASRGGLSSANLGH 2161
QY 1855 ---SHAQHPISPRTOALQORPSVLHNTGMGITAVPSKPTV 1897
D 2162 SAPPSSSQRTVGSGKSHLDGSSPEVKCSALDLP-----KSLYKGEKMTSS 2213
QY 1898 LRST---STSPVPAATPPATPCPLGLTD-GVYPTLMEVLLP-KEAPVAPPE- 1949
D 2214 SKGTGSAHSTAYVIGIKLT--PQVHNATPGELNISKIGFAPBSTVPSKDTVSYPQL 2271
QY 1950 ---RPRADTGHAFAPASGLEPPASPSKSGSPRLVPVSGH-ATVARTPAKMLAPH 2005
D 2272 HLRGQRSDRQHM---DPSQS-VKP--SPNEDGIKTLKLPQMGHRSIIEHIGSSRD 2325
QY 2006 HASPDPAAPASADPREKTQSKPFSIQELELSL-----GYHG-----S 2046
D 2326 RROKQKSSKETCKEKSSKSYLEPGVTTGEBGNLPERADEVLTFQFLGQPCNNVS 2385
QY 2047 SYSPGEVAVSPSPSLTHDKLPKHLE-----ELDKSHLEGELR---PKQGPVK 2095
D 2386 EKIGDKVLPLSGVPGKQSTQVEGSSKELOAPKCSKVATPLKMEGENOSKNTQESGP-- 2443
QY 2096 LGGEAALPHLRPLPESQSSSPLLOTPAVYKQHQRVTLAQNISEVITQDTHHQQOL 2155
D 2444 ---GSPAHIESVCPAPVSAKSP--GAGPGVQPSR-----NNTLSODPOSNNTQL 2490
QY 2156 ---SAPLPAVLPSF---PGASCVLDRRPPSLVLP-PPDHGAPARG--- 2196
D 2491 PEODRNLMTDGPPEQDGSFKRYPRSA---RASNMFGLTPLVGNRSYGEDI 2544
QY 2197 ---SPHSBGKRSPEPN----- 2210
D 2545 PYSNSGCKKRRKSAAGVQDADLSTSDLDLYYNTFTVTISGGEERLASHNLFRE 2604
QY 2211 ---KTVLGGEGCIGF-----PVSPPGCMTEPG 2235
D 2605 EEOCDLPKISQLDGVDDGSDTSVTATSRKSSQIPKRNKENGENTMLKIDRPEDAGEKE 2664
QY 2236 H-SRSAYVPLLYRDGEQTEP-----SRWGSKSPGNTSOPPAFFSKL-----TE 2277
D 2665 HVIKSAV-----GHKNEPKLDNCHSVSRV--KAQGDQSL-EMQLSLESRRVHTSTP 2714
QY 2278 SNSAVYSKKQELINKLNTNHRNPEYNIISQPGTEIF--NMPAITGTGLMTYRQAQOE 2334
D 2715 SDRKLLDTYNAELTKSDSDNNNSDCCNILLPSDLMDFLKTPTSMQALGESPESSSELL 2774
QY 2335 HASNMGLEALIRKALMGKTDOMEESPPLSANANPPLNANAS-----LPAAPPI- 2383
D 2775 TLGEGLDGLDS-NREKDIGLEVFESQOLP---ATEPVDSVSSSISAEQOFELELPSD 2829
QY 2384 ---TADGRSDHTLSPG----- 2398
D 2830 LSVLTTSPTVPSQNPRLAVISDSGEKRVITTEKSAVSSRGDALLSPGVDPAPBEGMT 2889
QY 2399 ---GGGKA---KVSGRPSSRKAKSP-----APGL 2421
D 2890 PDHFIOGHMADHISPPCGSVEQGHGNSODLTENSGPLQVPSPTVPVONOKYVSS 2949
QY 2422 ASGGRPSVSVSHSGCNRKRTPLTNR-----VMEDEPSSAG-----STPPY 2464
D 2950 TDSBPSPISAAVQTPPHLKPATEKLIIVNQNOMQPLVYIQTLPNGVTOKIQLTSPVSS 3009
QY 2465 NPLIMRLQAGVM-----ASPPPGI--PAGSGPLAGPHH 2496
```

```
D 3010 TPEVMENNTSVLAPMMSGULTTGTGNPSLPPSPSLPPAPAKGLLSVPHH 3058
RESULT 35
JC5630
TCOF1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 05-Nov-1999
C:Accession: JC5630
R:Patnckas, W.A.; Zhang, N.; Gridley, T.; Jabr, E.W.
Biochem. Biophys. Res. Commun. 238, 1-6, 1997
A:Title: Mouse TCOF1 is expressed widely, has motifs conserved in nucleolar phosphoprotei
A:Reference number: JC5630; MUID:97445113; PMID:9299440
A:Accession: JC5630
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1320 <PAZ>
A:Cross-references: DDBJ:AF001794; NID:92109458; PIDD:AB71347.1; PID:92109459
C:Comment: This protein is a nucleolar phosphoprotein with 82 potential phosphorylation s
ing.
Query Match 2.3%; Score 389; DB 2; Length 1320;
Best Local Similarity 20.4%; Pred. No. 1.2e-06;
Matches 354; Conservative 193; Mismatches 660; Indels 530; Gaps 75;
QY 479 YKSLVRSYRRRKSGQQQQQQQQQQQQ-----QQQPMRSGSEDEKEKEAE 530
D 15 YHHLQGYRAAREVEQSGKSFLLQPVTLIDITTHWQOTSELQKQAEDETLQAK 74
QY 531 KESEKPEVNDKEDLLEKTDPTSGEDNDEKAVASGKRTANSQGRKGRITRSMANEA 590
D 75 KSRVSDPVSSSE-----SSDQKEBEALTEAKATPR-----PTPV 110
QY 591 NSEBAITPQGSABELASMEINSSRWTEEMETAKGLLEHGRMSAIAKRVGSKTSYOCK 650
D 111 NSATAALPSKYKE-----KGRKTKANKTV-----NSVSHSGSKTV----- 146
QY 651 NPTFNKYKRONDELILQOHLKMEKENARKKKKAPAAASEAPFPVVEDEMEMASGV 710
D 147 -----VHLL-----SGSKPKSKAEPLANTVLA----- 168
QY 711 SGNBEWVEBAEALHAGNEVPRGECGPAIVNNSPTESIPSPHTEAAKDTGONGPKPP 770
D 169 SETEBEAGNALPTAKSGTVSAGQS-----SSSDSSISDET-----DVEYKSAKP 218
QY 771 ATGADPPPGPPTPPRTSRAP-----IEPTPASEATGAATPPAPSPSPAPPPVPE 825
D 219 AQAKASAP--AKOPPARTAPGPTKGNVAPTPAKPAAAAAANAANAANAEESE 276
QY 826 EKEBETAAAPVEBEGEQPPAAEBELAVDTGKAEPEVKSCTEBAEGCPAKGDAEAAEA 885
D 277 SSEEDS-----DSEBEAPAGLPSQVYKASG-----GPHVRADSVASAG 314
QY 886 -TAEGALKAEKKGSGSRATTTAKSSGAPQODSSATCSADEVDABEGDKNRLSPRSL 944
D 315 ISGKPIPLAPKGTGR-ATQAKAREPKDSSTSSSDSSSEDE-----M 358
QY 945 LPTGDRANASFOKPLDKOLKORAAAIPIQVTKVHEPRDEDAFTKAPAPPPPN 1004
D 359 PVTNTPQARTSGSP-----RA-----RGSAPAK----- 384
QY 1005 LQPSDAPQPGSSPRKRSPPAPADK-EAPAA-EAQKLPQDPBPCTSGLPFPVPPREV 1062
D 385 ---ESSQKAPAVTP-GKAPRVAAQAKPKAKSSSESSDGETPA----- 426
QY 1063 IKASPHAPDSASFYAPPGPLPLGLHDTARPLYLPPTTISN-----PPPLISSAKHP 1115
D 427 -----ATUTTSAPVYKPLGKSSQVPRVSTYTPGSSGGGANLPCGKXGSAALRV 476
QY 1116 SVLERQIGALSQMSVOLHVPYSEHAKAPVGYTMGLPLMDPKCLAP----- 1163
D 477 QWYKE--DVSSESSAEL-----DSDGPGSPAKAKASIAL-POKVPVATQVKTDRGKG 527
```


R. Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A:Description: Fox genes: pollen-specific genes with extensin-like domains.

A:Reference number: S49915

A:Accession: S49915

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1188 <RUB>

A:Cross-references: EMBL:Z3465; NID:G600117; PIDN:CAA84230.1; PID:G600118

Query Match 2.9%; Score 389.5; DB 2; Length 1188;

Best Local Similarity 18.6%; Pred. No. 1.1e-06;

Matches 292; Conservative 104; Mismatches 369; Indels 807; Gaps 64;

```

QY 733 RGECSRPATVNNSSDTEISIP-----SPHTAADTQNGPKP---PATLGADPPRG 781
DB 374 RNCICGLALPRAOKTALQCAFLVLRVDCSHVCAGYPTPGCGSPSSPVKPKAASPMBS 433
QY 782 PPTPRRTSRARLEPPTASATGATPTTPRPA--PSGAPRPVVKKEKEETAAAPVEES 840
DB 434 PHTPP-----DVSPRLPEPSFVPRAPAPMPTPHSP-----A 467
QY 841 EEOKPRPAEELAVDTCKAEKRVKSECTEAEKSPAKGDAEAAATAGALKAEKGGG 900
DB 468 DDVVPPTP-----PVPKSPRA-----484
QY 901 GRATTAKSGARODSSSATSCADEVDEAEKGDKRNLSPRSL--LTPTGDRANASPO 958
DB 485 -----TSPSPQVQBPAAAS-----TPPSELVTKSP-----1010
QY 959 KPLDQLQKQRAAATRIQVTKNRPREDARTKPA--PAPRPQNLQRESDAPOQ 1015
DB 511 APV-----GSPRPVKT--SPARTGSPSPRPVSVSPRPVKSPP--PPAP 555
QY 1016 GSSPRGSKSPAPRAPDKEAFAAEAOQLPGDPCMTSGLPFPVPREVIKASPHAPPSAF 1075
DB 556 VGSPPRPKSPPPRA-----PVASPP-----PVKSPRPPTLVAS--PPRPVK 596
QY 1076 STAPRPHPLRLGHTARPLPPTTISNRPPLISSAKHPSVLERQIGALSQMSVOLAH 1135
DB 597 SPPPRA-----PVASPPRPKSPRPPTPVASPP-----624
QY 1136 PVSENAKAPVPTWGLPLPMDPKLAPFGVQKQEOQLSPRGQAPRESIGVPTAOEASVL 1195
DB 625 -----PPAPVASSPP--PMKSP-----PP-----642
QY 1196 RGTALGSLVPGSITKGIPTSTRVPSDAITYRGSITHTPADVLKGTITRIIGEDSPSRL 1255
DB 643 -----TPV-----645
QY 1256 DKGREDSLPKGHVITGCKGHVLSYEGGMSVTQCSKEDGSSGPRPHETAPAKTIDME 1315
DB 646 -----SSPP-----649
QY 1316 GRVGRALISSASIEGLMGRAIPRRNHSRPHLKEQHNRGSIITQGISYVAQEDYLRRRA 1375
DB 650 -----PEPK-----653
QY 1376 KLIKREGTPRRPSPSRL--TEAYKTQALRLKRAHEGLVATVXKAGSINHEIREEL 1433
DB 654 -----SPPPPPASTPPEEYPTP--PTSVKSSP-----PPKKS 686
QY 1434 RHTPELPLARPLKEGSIITQGTPLKDTGASTGSKKHNVRLSIGRGRTFPPVHLDVMA 1493
DB 687 LPPTPLIPSPRP--QEKPTPSTPSKSPS-----SPKSPSP-----721
QY 1494 ADBALERACYSRSLSRPGLTASSGGSIARGAPVIVPELIGKRPQSLPTEDHGAAPAGH 1553
DB 722 -----KPPVSSPPTPKSPPPAPVSSPPTVSSPPALAPVS---SPSPVKS 766
QY 1554 LPRGSEVTAKKEPRLQOEGSLSSSKASODKLTSTPRELAKSPHSTVPRHNHPRISPEYH 1613
DB 767 SPBPATLSSPPAPVQK-----SSPPPVQVSSPPAPKSSP--PLAP---806

```

```

QY 1614 LARGVGVULYKSHITPLADPTSIERGFLDAAAYVLRPHLAPNTYPHLY---PPYL 1670
DB 807 -----VSS-----PPVEKTS--PPAPLSSPP-----LAPKSPHHVVVSSPPPV 846
QY 1671 RGVPTPALLENQTIINDYITSCQMHNTATAMAQRADMLRGLSPRESSLALNYAAGPRG 1730
DB 847 KSSPPAPVSSPP-----LTKRAPPAPVSSPP-----876
QY 1731 IIDLSQVPHLPVLPPTPTPATMDRLAYLTPAPQPFSSRHSSSPSLFGCPTHLTKPPT 1790
DB 877 -----VKESTPPAPPTVISP-----PSEPK-----SSP--PPTVSLPPTIV 912
QY 1791 TSSSERERDRDRDRDREREKSLTSTTVNHARITWRGTQSSSGSGSGGSGSSSR 1850
DB 913 KSSPP-----917
QY 1851 PASHSHAHQHSPISPRTDALQORPSVLNHTGMKGIITAVBSKPTVLKSTS---TSSP 1906
DB 918 PAMVS-----SPMTPKSSP-----PPV-----VSSPPPTVKSPPAPVSSPP 957
QY 1907 VBPATFPPTATCPRLGSLIDGVYPTLMEVLLPKKARVAPRBRPRADTGHAFLAKPPAR 1966
DB 958 ATPKSSPPRA-----PVNLPPREVKSSPPTPVSS-----SPPPA--991
QY 1967 SGLERASSPSKSGSPRPVLPVPSGHATITARTAKNLARHNASDRPARASADPRREKT 2026
DB 992 -----PKSSP--PPAPMSSP-----PREVKSPPPAPVSSPPPP--V 1025
QY 2027 QSKPFSIQLELRLSLGYHSSVSPGCEVPSVSPSPSLTHDKGLPCHLELDKSHLEGEL 2086
DB 1026 KSP-----PPAVSSP-----1037
QY 2087 RPKQRPVULGEBAAHLPHLRPLPSOSPSPLLQTPAGVKGHQRVTLAQLISEVITOD 2146
DB 1038 -----PPVKSPPPPAPVSSPPPPVKSPPPPADISSPPPVK-----1074
QY 2147 YTRHNRQQLSAPRLAPVLPFPGASCVDLRRRPSDLYLPRRDHGAPARGSPHSEGGKS 2206
DB 1075 -----SPPPPAVSSPP--PPVKSPPPPAPVSSPPPPVKSPPPPVSS--S 1116
QY 2207 PERNKTSVLGGEDGIEPVSPREGMTPEGHSRS--AVYPLLYRDEQTEPSRMGSKSP 2263
DB 1117 PPP-----APVKPP--SLPPAPVSSPPVTVTPAPKKEQSLP-----PP 1155
QY 2264 NTSQPPAPFSKL 2275
DB 1156 AESQPPSPFNDI 1167

RESULT 34
A48205
A11-1 protein +GTE form - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 22-Jun-2003
C/Accession: A48205; B48205
R/Name, Q, J, Alter, H, Nelson, K.K.; Chatterjee, D.; Gu, Y.; Nakamura, T.; Canaan, E.; Crc
Proc. Natl. Acad. Sci. U.S.A. 90, 6350-6354, 1993
A/Title: Analysis of the murine A11-1 gene reveals conserved domains with human A11-1 and
A/Reference number: A48205; WUID:93317679; PMID:8327517
A/Accession: A48205
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1505, 1509-3869 <HA2>
A/Cross-references: GB:L17069
A/Accession: B48205
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1505, 1509-3869 <HA2>
A/Cross-references: GB:L17069
A/Genetics:
A/Genes: A11-1
C/Superfamily: histone methyltransferase, trithorax protein type

```


Db 843 GCHAPFMSTVPPRL-----PPASIGAAVQPKMESGT-----PAGPPENVLPISM 890
Qy 1217 VPSDSAITYRGSIITHTPADVLKGTITRLLIGDSFRLDRGEDSLPKGHVIEGSKGH 1276
Db 891 APPLSL-----GLRGHGP-----QTEPTKVE-----VKPVP----- 917
Qy 1277 VLSYEGMSTVQCKEDGRSSGPPHETAPKRTYDMERVGRAISSAIEGLMRAIP 1336
Db 918 -----ASPPKPKVSAVVSPOW-----KALACVSAEGVVEEPA 952
Qy 1337 PERHSHHLKEQHNGIRGSIQTGIPRSYVEAQEDYLREAKLIREGTPPPPSRDLTEA 1396
Db 953 SEHLKP-----ETQETPREKPRPLRYTKAVPT-----RQST-V 985
Qy 1397 YKTQALGPKLKPAGHGLVATVKEAGSLHEIPREBLHTPELPLAPRLKESITQGT 1456
Db 986 PKLPAVHPARLR-----KLSFLPTPR-----TQGSSE 1011
Qy 1457 LKVDTSASTGSKKHVRSII-----GSPQRTPPVPHPLDVMADARALEACYEBSLSKR 1511
Db 1012 DVQVAFISETIGIASLSLLEQFEKSEAKKECPAPADSLA----- 1054
Qy 1512 PGTAASSGSGSIAGAPVIVPELGKP-----RQSPITYEDHGAPFAGHLPRGSPVTMRREPTPR 1568
Db 1055 -----VGNSSG-----VDIFQEKRPDLRLQAP-----ELANVAGLTP---PAT---PRHQ 1093
Qy 1569 LQEGSLSSSKASQDRKLTSTPREIANSPIHSTVBNHPHPISPYBHLRGVSGVD----- 1622
Db 1094 LMKPPLAVALSTAKAKPKSTAQGTLPKEGVTEAKHPAAV-----RLQEGVHGSRAVHGS 1149
Qy 1623 -----LYRSHIPLAPPTSPRGIPIDAAAYVLPFR-----LAPNTYHLLVPPYL 1669
Db 1150 GMDHYCVRSRTP---PKKMPALVIVPVGSRMVKKHQDITIKVSLGAAAP---PP-- 1200
Qy 1670 IRGYPPTAALENQTIINDYITISQOMHNATMAQADMLRGISPRESSIALNTVAAPR 1729
Db 1201 -----PCIAA--SREPL--DHRTSSEQADPSAPCLAPBS---LISP-EASPCAN----- 1241
Qy 1730 GIILDSOVPLPVLVPTPTPTATAMD--RLAYVLPAP-----QPPSSRHSSPLSPGCP 1782
Db 1242 -----DMNTRTPPEPAKQSRMRCYKACRSAPSSQGWGRGRSRSVSSGSR 1292
Qy 1783 THLTCTTTSSSRERDRDRERDRERKSIITSTTVHAPITWPGTQSSG----- 1836
Db 1293 TSEBASSSSSSSSSSSRSLSPPHKR-----WRSSCSSSGSRNRCS 1336
Qy 1837 SSGSGGGGGSSSRPASHAHQHSPISPRTODALOORPS 1876
Db 1337 SSSSSSSSSSSSSSSSSSRSRSP-SPKRRSDRRRRTYS 1375

RESULT 31
A41948
alpha-fetoprotein enhancer-binding protein - human
N/Alternate names: ATBPI protein
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C/Accession: A41948
R/Morinaga, T.; Yasuda, H.; Hashimoto, T.; Higashio, K.; Tamaki, T.
Mol. Cell. Biol. 11, 6041-6049, 1991
A/Title: A human alpha-fetoprotein enhancer-binding protein, ATBPI, contains four homeod
A/Reference number: A41948; MIM:32049333; PMID:1719379
A/Accession: A41948
A/Molecule type: mRNA
A/Residues: 1-2783 <MOR>
A/Cross-references: GB:D10250; GB:D90395; NID:G219429; PIDN:BA01095.1; PID:G219430
A/Note: sequence extracted from NCBI backbone (NCBI:66271, NCBI:66276)
C/Genetics:
A/gene: GDB:ATBPI
A/Cross-references: GDB:392090; OMIM:104155
A/Map position: 16q22.3-16q23.1
C/Superfamily: alpha-fetoprotein enhancer-binding protein; homeobox homology

C/Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger
F,128-94/Region: zinc finger CCHH motif
F,128-150/Region: zinc finger CCHH motif
F,176-198/Region: zinc finger CCHH motif
F,311-332/Region: zinc finger CCHH motif
F,340-361/Region: zinc finger CCHH motif
F,448-471/Region: zinc finger CCHH motif
F,489-509/Region: zinc finger CCHH motif
F,517-538/Region: zinc finger CCHH motif
F,633-655/Region: zinc finger CCHH motif
F,684-706/Region: zinc finger CCHH motif
F,719-773/Region: serine/threonine-rich
F,809-958/Region: glutamine-rich
F,1071-1092/Region: zinc finger CCHH motif
F,1117-1211/Region: proline-rich
F,1232-1288/Domain: homeobox homology <HOX1>
F,1329-1385/Domain: homeobox homology <HOX2>
F,1416-1437/Region: zinc finger CCHH motif
F,1618-1638/Region: zinc finger CCHH motif
F,1728-1784/Domain: homeobox homology <HOX3>
F,1799-1820/Region: zinc finger CCHH motif
F,2033-2089/Domain: homeobox homology <HOX4>
F,2112-2134/Region: zinc finger CCHH motif
F,2545-2566/Region: zinc finger CCHH motif
F,2585-2607/Region: glycine-rich
F,2611-2633/Region: zinc finger CCHH motif
F,2650-2737/Region: serine/threonine-rich

Query Match 3.0%; Score 394; DB 1; Length 2783;
Best Local Similarity 18.2%; Pred. No.1.9e-06;
Matches 510; Conservative 273; Mismatches 953; Indels 1064; Gaps 115;

Qy 99 EMFTSKPRLELDPLRPEPLATGQPSGEBLT-----KDRSLTGLEVPSPS-- 152
Db 535 ETSHELEADIDQILGGLANGDLAMGDPPLAEHHTIIVEDKEESDLEKQSTGS 594
Qy 153 -----PPTDPELVLPVPR-----LSKEELIOMNDVDRITWV 186
Db 595 DSGSVQEDSGSEPKRALPPKGPNTMEKFLDPSRYKCTVCKESTQ-----KILIV 648
Qy 187 E-QQISLKKQOOLEEAAKPEPEKPVSPPIESKHSLVQIIVDENRKAAMARIL 245
Db 649 HYNVSVHLMKRLALDESATGQEPETSPDNKPFKNTCNVA----- 690
Qy 246 EGIQVVELPLNQSDTFOYENIKINDAMRKLLILYFKRRNHAKQWKQRCQYDL 305
Db 691 -----YSQ--SSTLEIHMRSVYLQV----- 708
Qy 306 MEALEKKVERIENPPRRAKESVREYERKQPEIRKQELQERMSQVQSGSLMSMA 365
Db 709 -KAAKLEAAGSSNGTNGSSSI-----SLSSSTSPVSTSGSNFTTIS 782
Qy 366 ARSEHEVSEIIDGLSEQNLKQKQALAV-IPMLYADAOQRIKFTNMGLMADPMKVYK 424
Db 753 NPSS-----AGIASSNLISQVPTESVGMP-----LGNPICA-- 785
Qy 425 DROWNMMSBOE-KETFRKFMQHPKNGILASFLEKTYACVLYYYLTKNENKSLV 483
Db 786 -----NIASPEKPEKANKRLAD-----MIS- 807
Qy 484 RSEYRRGKQQQQQQQQQQQQQQQQQQPPRSGQEEKD--EKEKEAKEEKEE- 535
Db 808 -----RQQQQQQQQQQQQQQQQQQQQQQLAQAQVQAHQDELQQAALIGQLRPPL 862
Qy 536 -PEVENDKEDLLLEKTD-----DTSGEDNDEKAAVASKRTYANSQGRKRRITR 584
Db 863 LPIFPMTTEILLQOQQOHLFPFYIPSAFQNPVSLPVTSGALLTGTG--PGLLD 920
Qy 585 SWAN---EANSSEAITPQOSAEIAS-----MEINSSRWTEEME 621
Db 921 LKAQVQVPOOSHQOILPQOQONOLSTAGSHSALLQPSQHEKKKKLVIREKEKESQERD 980
Qy 622 TAKKGLLEGHGMNSAIARAVGSKTVSQCKNFYNYKKQQLDEILLQOHLKMEKENARR 681


```

QY 1467 GSKKHIVRSILGSPRTFPPVHP-----LDVMAADARALERACYEESIKSRPGTAS 1516
Db 900 -----IGPPG---PBNCTPGHRSFGIPGVKGGRTGTGAGGEGDCKNPNPSE 946
QY 1517 SSGGSIARAGAVIYVPELGKPRQSPFLYEDHQAAPAGHLR- RGSPTVMTREPTRELQSGSL 1574
Db 947 ISHVIGDKGEPGLKGFAGNPGGEK-----GNRGVPMPPGLKGLPGAPGPPGR---GDL 999
QY 1575 SSSKASDRKLTSTPRRI-----AKSPHSTV-PEHHHPISPYEHLRQVSGVULYR 1625
Db 1000 GSTGNPEPGLIGPGSMGMGMGSKGKGTGLGFPAGRPGGLPGHIGQGDGEBGYS 1059
QY 1626 SHI-PLAFDPTSIIRGIPLDAAAYVL-----PEHLAPNPTPYLHYPPYLLRGVPTAAL 1679
Db 1060 EGRPRGPGFTGDP-GLRPGMGKKGEMGQRPBHL--GAPGPGAR-----GSPGSPGL 1111
QY 1680 ENRQTLINDYITSGQMHNTATAMAORADMLRGLSPRESSIALNVAAGPGLIIDLQVPH 1739
Db 1112 PGK-----PGRHDLGFKGKIG 1128
QY 1740 LPLVLP-----TCTPATMDRLAYLPTAROPSSSRSSPFLSGPHTLTKPTTSSSE 1795
Db 1129 L--LGPPGIRGPPGLPG-----FGSPGPMGIR----- 1154
QY 1796 RERDRERDREREREKSLTSTTVEHAPIMRPGTESGSSGSGSGG---GSSSR 1850
Db 1155 --GDQHGDIIPGAGEK---GETGLRAB--PGRGNPDAQAKDRGAPGPGPLGR 1205
QY 1851 PASHSHAHQSPISPTOTALQORPSVLANMGKIIITAVEPSKPYLRST--STSSPV 1908
Db 1206 KGMAGDA-----GPRPGTIGBPP-----GPRGLPGAILPGQ-TGNRGPGRSGPGA 1252
QY 1909 PAATTPPATCPGGLTIDGYITLMEVLLPKAARVAREPRADTGNHFLAKPRRSG 1968
Db 1253 PCGPPGPGSHV-IG--TKGDKSGWGHG--PKGPPGTA-----GDMG--PGRIG 1295
QY 1969 LEPASSSKSGSEPRPLVPVSGHATARTPAKNLARHNASDPRAPASADPHREKTQS 2028
Db 1296 -----APGTGLDPRGDP-----GP 1311
QY 2029 KPFSIOLELRLSLGHSYSYSPGVEPVSFVSPSLTHDKLPHLIELDKSHLEGELRP 2088
Db 1312 QGPPGVKGEKGNPGFLG-SIGPPG--PIGKPPGVAGDPTLKIIS-----LP 1357
QY 2089 KQGPVYKGEBAHLPHLRPLPSQSSSPULLQTAPEVKGHQRYVTLAQHI SEVITODYT 2148
Db 1358 GSPGPGTPEG-----PGMGSE----- 1374
QY 2149 RHHPQOLASPLRPLVSPGASCVLDLRPRSDLYLRPPDHGAPAR-----GSP 2198
Db 1375 -----PG---PG-----PGLNG-PCGPRGKPGGDKGRTGPPAGEK 1408
QY 2199 HSEGGKSPENPTSVLG-----GEDGLEPVSPEGTEPGRSRAVVP----- 2243
Db 1409 GNGSGEPBPASDGLRGLKGRGDSGPAITTRTFVTRHSQTLAISCEBGTVPV 1468
QY 2244 ----LYRDEQ-----TEBRMGSKSPGNTSOPRAPFSKULTESNASAVKSKOEINKLN 2295
Db 1469 SGFSFLFVQGNQRAHGODLGTLSCLQRTTMTPLFLFCN-----VNDVCN 1512
QY 2296 THNRNEPVNI SQPTEIFNMRAITGTGLMTYSQ-----AVQEHASTMNGLEAIT 2346
Db 1513 FARRNDISYVLTSTPALMPMMAPIITGALPYSRCTVCEGPAIALAVHSQTT----- 1565
QY 2347 RKALMGKYDQWESSPLSANAFNPLNASASLPAAMP---ITAADGRSDHTLTSPGGGGA 2403
Db 1566 -----DIP-----CHGWISLWKGBFIMFTAGSGGTG 1595
QY 2404 KYSGRPSRRAKSPAPGLASGDRPPSVSVHSGDCN 2440
Db 1596 QALASPOS-----CLEEFRASPLFLECHGRGTGN 1623

```

```

RESULT 30
T00273
hypothetical protein KIAA0595 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C/Accession: T00273
R/Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Koreani, H.; Nomura, N.; Ohara, O.
DNA Res. 5: 31-39, 1998
A/Title: Prediction of the coding sequences of unidentified human genes. IX. The complet
A/Reference number: Z14086; MUID:98290545; PMID:9628581
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1520 <NAG>
A/Cross-references: EMBL:AB011167; NID:g3043713; PIDD:BA025521.1; PID:g3043714
A/Experimental source: Brain
C/Genetics:
A/Note: KIAA0595

```

Query Match 3.0%; Score 395; DB 2; Length 1520;

Best Local Similarity 22.4%; Pred. No. 8.9e-07; Matches 331; Conservative 148; Mismatches 515; Indels 486; Gaps 76;

```

QY 528 EAEKEEKPEVE--NDKEDL--LKEKTDDTS-----GEDNDEKAVASKGRKTANSQ 575
Db 251 ESETEAVPKVTLCSKEGSLNSSEKLDASCLIKREVEVEPVYKPEPPQPNANAAGSQ 310
QY 576 GRRKRITRSMANEANSEAITPQASALASMBLNESSRWTEEMETAKKGLLEHGRWS 635
Db 311 RARGRKKKSEKQ-----PAACVEGYARRLSSSR-----GQS-- 343
QY 636 AIAKMSKTVSCQKNFYRYKKRQNDLQGHKLKMEKRNARRKKKKAPAAASEEA 695
Db 344 ----TVGTEVTSQ-----VDNLQKQPEELQKESGELQKGRPRAMARAWAA 386
QY 696 FPPVDEDEMEASGVSGNEEWEAEALHAGSGNEVRCGSGPATV--NNSSDT--ESIPS 753
Db 387 -----ALENSSPKULE-----RSAGSSSAKE--GPLDIYPLADITQINPI 426
QY 754 PHTAAKDTQNGKPKPATLGADGPPGP-----PTPP-----RTTSRAPIETPASEA 802
Db 427 PTHLSLVDSAQASBPMVDSEADPTAVGAVLPGVPDGLVDLASTSSSLVEPLPAEPV 486
QY 803 TGAPTPPAPPSAPPPVYPKKEKEETAAAPVEEGEQQKPPAAELAVDTGKAEPV 862
Db 487 L--INPVLADSAAVPAVPAVIDN-----LPPVD--AVPSGPAVDIAL-----VDPV 530
QY 863 KSECT-----FEAEKGPAGKDAEAEATA-----EGALKAER 895
Db 531 PNDLTPVDVPLVAKSRPTDPRKGAIVSSALGSAFQLVSESLDPPKTIIEVKEVYDLSK 590
QY 896 KEGSGRATTAKSGAPQDSASATCSADEVDEAGDKNRLSPRSLLTPTGDPANA 955
Db 591 IESGTS-ATTHEA--RRRPLSLSEYRRRRQQAETEE--SPQ-----PTG--KMPG 637
QY 956 SPQKPLDKQLKQRAAIPPIQYTKV-----HEPPREDA-APTKAPAPAPPP----- 1001
Db 638 LPETPLGLADIP--CLVTPPAKPAKTALQRPETPLEICLVPGVSPSPASPEBPVSKPV 695
QY 1002 ----PQNLQPSDAPQOGSSPRGKSRSPAPPAUKKAAABAQULPGDPCCMTSGLPFV- 1057
Db 696 ASSPTQOVPSQEMPLARBPSPVQSVSPAVPT-----PESMSAALPPFAG 740
QY 1058 -----PREVIKASPHADPSPAFSVP- PGHPLPLGLHTARPVLPPTTI 1102
Db 741 GLGMPSSLPPPLQPSLPLSMGVLPDPPT-HYALPLSPRCVPHVSPSGYPCLPPTPV 799
QY 1103 SNPPPLISSAKHPSVLERQIGALISQGNVOLHVPYSEHAKAPVGP-----VTMGLPLMDP 1158
Db 800 ----PLVSGTP-----GAYA--VPPTCVPMAP-PPAPVSPYSSTCTYG-PLQWGP 842
QY 1159 -KTLAP-SGVKQEQLSPRQAGPPESLGVPTAQEASVLKGTALSGVSGSITKGIPISTR 1216

```

RESULT 29

CGH03B

collagen alpha 3 (IV) chain precursor, long splice form - human

N:Alternate names: Goodpasture antigen; procollagen alpha 3 (IV) chain long splice form

C:Species: Homo sapiens (man)

C:Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999

C:Accession: A54763; A43928; A44043; A45971; A39786

R:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Trygvaason, K.; Reeders, S.T.

J. Biol. Chem. 269, 23013-23017, 1994

A:Title: Complete primary structure of the human alpha3 (IV) collagen chain. Coexpression

A:Reference number: A54763; MUID:94364994; PMID:8083201

A:Accession: A54763

A:Molecule type: mRNA

A:Residues: 1-1670 <MAR>

A:Cross-references: GB:X80031; NID:G577563; PID:G577564

A:Experimental source: Kidney

R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.

J. Clin. Invest. 89, 592-601, 1992

A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the al

A:Reference number: A43928; MUID:92147878; PMID:1737849

A:Accession: A43928

A:Molecule type: mRNA

A:Residues: 1331-1524, 1', 1526-1670 <TUR>

A:Cross-references: GB:M81379

A:Experimental source: Kidney

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 267, 19780-19784, 1992

A:Title: Exon/intron structure of the human alpha 3 (IV) gene encompassing the Goodpastur

A:Reference number: A44043; MUID:93015826; PMID:1400291

A:Accession: A44043

A:Molecule type: DNA; mRNA

A:Residues: 1386-1670 <CUI>

A:Cross-references: GB:M92993; NID:G177895; PID:AAA21610.1; PID:G177896

A:Note: sequence extracted from NCBI backbone (NCBI:P115587)

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 269, 17358, 1994

A:Reference number: A44738; MUID:94274734; PMID:8006044

A:Contents: annotation; exaltum; correction to intronic sequence in A44043

R:Bernal, D.; Quinones, S.; Saus, J.

J. Biol. Chem. 268, 12090-12094, 1993

A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.

A:Reference number: A45971; MUID:93280184; PMID:8505332

A:Accession: A45971

A:Molecule type: mRNA

A:Residues: 1427-1444 <BER>

A:Note: sequence extracted from NCBI backbone (NCBI:P133363); sequence incorrectly ident

R:Morrison, K.E.; Maruyama, M.; Yang-Feng, T.L.; Reeders, S.T.

Am. J. Hum. Genet. 49, 545-554, 1991

A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of

A:Reference number: A39786; MUID:91355570; PMID:1882840

A:Accession: A39786

A:Molecule type: mRNA

A:Residues: 1453-1593, 'A', 1595-1670 <MOR>

A:Cross-references: GB:555790; NID:G234418; PID:AA19637.1; PID:G234419

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (

ed and subsequently O-glycosylated.

C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope

C:Gene: COL4A3

A:Gene: GDB:COL4A3

A:Map position: 2q36-2q37

A:Residues: 1385/1, 1418/1, 1488/1, 1547/2, 1585/3, 1643/2 #status incomplete

A:Note: The alpha 3 (IV) and alpha 4 (IV) chain genes are encoded on opposite strands with

C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3

er associations in the interrupted helical domain (with disulfide and desmosine cross-l

C:Description: minor structural component of extracellular basement membrane in kidney

C:Superfamily: collagen alpha 1 (IV) chain

C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1670/Product: collagen alpha 3 (IV) chain, long splice form #status predicted <MAT>

F:29-42/Domain: amino-terminal nonhelical, NH1 <NH1>

F:43-1438/Region: interrupted helical

F:791-793/Region: cell attachment (R-G-D) motif

F:996-998/Region: cell attachment (R-G-D) motif

F:1154-1156/Region: cell attachment (R-G-D) motif

F:1306-1308/Region: cell attachment (R-G-D) motif

F:1345-1347/Region: cell attachment (R-G-D) motif

F:1432-1434/Region: cell attachment (R-G-D) motif

F:1439-1670/Domain: carboxyl-terminal nonhelical, NCL <NCL>

F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>

F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>

F:31,33,39,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status predi

F:253/Binding site: carboxylate (Asn) (covalent) #status predicted

F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted

F:1505-1511,1616-1622/Disulfide bonds: #status predicted

F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 3.0%; Score 397; DB 1; Length 1670;

Best Local Similarity 21.1%; Pred. No. 8,4e-07;

Matches 388; Conservative 121; Mismatches 624; Indels 704; Gaps 91;

QY 752 PSHTFAKKTQNG-PKPRATIGADGP-----PPRPPTPRRTSR-----AATLP----- 796

Db 343 PTEYTDYDEKGBGTPGPPRGARGPQSPGPVGPSP-SSSRGLGAGCWPLIG 401

QY 797 -----TPSEATGATPPRPPSPSPAP-----PPVVKKEKEBETAAPVEEGEEOKPP 846

Db 402 SKGERGAPKQANGTQSPSCAGSPGLPSGPP-----GPGDIDGFRKGP 448

QY 847 AABEL-----AVDTGKAEVYKSECTE-EAERGPAKGDAAEAATAGALKAEK 895

Db 449 GDHGLPEYLSPGLPGVDGKGPGLL-CTQCPYIPGP-----PGLPGLPLHG 496

QY 896 KEGSGRATTAKSGAPQDSATCSADVIDEABG-----GKNLLSRPSLLTPTGD 950

Db 497 VKGIPIRGQAAGKSGSPSGNGLPQPPGPAQGPGLKSGKGETLQPEGGVAVP-GD 555

QY 951 PRANASP-OKPLD-----LKQKORAAIPIPLQYTKVHEPRREDAPFKPAPRAPP 1002

Db 556 PGLRGPGRKGLDGTRETPEVKGLPPKPKGLALSGEKGDGPPDPSPSPGAPGAP 615

QY 1003 QNLQESDAPQOGSSBRGKSRPAPPADKEAPAEAKOKLGPDPCTWSGLPFVPPREV 1062

Db 616 -----POYGPQ-----EPGLQGVGVAP-----GPPGAPGRGE 647

QY 1063 IKASPAAPDSAPSVAPRPHPLPLGLHDTARPLPRPPTISNPPRLISSAKHSVLERQI 1122

Db 648 LSVSTPVPQPPG-PGPPPHHGFQ-----PPGI-----PGLSGKGDGL 687

QY 1123 GALSQGSVGLHVPYSEHAKAPVGTWGLPLPMDPKLAPFSGVQEOQLSP--RQOAGP 1180

Db 688 -----PPDGGEPGIP-----GIGPPGPPGKGGGFGTGKSLCPGKMEBGL 731

QY 1181 PESLGPPTAQEASVLKGTALGVPQGSITKGIPTSTRVPSDAITYRGS-----THGT 1233

Db 732 PGRPGLPGA-----KGRPAVAMPDGGTGTGPFGEKGNSE-----HGEIGLPLGLPGLT 781

QY 1234 PADLVYKGTITRILIGEDSPRLDRGBDLSLPGHVIYEEKKGVLSYEGMSVYQCKED 1293

Db 782 PGNE-----GLDGP-RGDPGQ----- 796

QY 1294 GRSSGPRHETAPKTYDMMEGRVGRATSSAIEGIMGALIPRHRSHPHLKEONHTRG 1353

Db 797 -----PGPPGQPPGKRIEPPRQAQ-----LPGLNG-----LNGQGGRRG 833

QY 1354 SI-----TQGIPIRSYVEAOEDYLRRKAKLKRSESTPPPPSRDLTEAYKTOALGPLTKP 1409

Db 834 KTRPKDPPGIPG-----LDRSFPG-----ELGSPG-----IP 861

QY 1410 AHGELVATVKEAGRSIHETPRELRTPELP-LAPPLKEGSI-TQGTPLKTDGASTT 1466

Db 862 GHQEGMPLQGRG-----YPGNPILGP-PGEDGVIGMMGP-----GA--- 899

```

Db      679 DDPAKPVSLB-----PEKTPVLAKAP-----TKPDSE 710
Qy      1705 QRAMLGLSPRESLALNTAAGRGIIIDLSQVHLPLVLPPTGTATAMDRLAYLPTA 1764
Db      711 AAAPVSGPSSKDKLAKKAPVKPRDPSPMKAVPIKP--AKTEVPAAV-----KK 760
Qy      1765 POPSSRRSSSP-----LSPGPTLTKPTTSSSERERDRDRDREREKSIITSTTT 1820
Db      761 PEPAKGRDPSPKAKKAPNSP--VPPPTPVKAPVK-----795
Qy      1821 VEHAPIRPGTEQSSSGSSSGSSGSSRPASHAHQHSPISPRTODALQORPSVLHN 1880
Db      796 -----MKPWEDD-----AAEPVNVPEPEKTPVLAKTPVKPRDPS-----834
Qy      1881 TGMGIIITAVPSKPTVLRSTSTSSPVAPATPPATHCPLGLDGVYPTLMPPVLLPK 1940
Db      835 -PKAV-----PAKP-----STKTDAF--PVSVKKP-----EPVSKPK 864
Qy      1941 E-APRVAAPERPAD-----TGAFLAKPPARSGLEPASSPSGSEPPRLVPVSGHATIA 1995
Db      865 EPSKAAEPNSPVVPPPVKAPVKKKKPPWEDDEPTEVKKPSPEKTPVLAKKEPEK 924
Qy      1996 RTPAKNLAFPHASDPAPASASDPAREKTOQKPFISOELRLSLGYHSSYSPEGV-- 2053
Db      925 PKDAPKVA--AKRDRPEPKKAVPEKEPAKVAAKPRDLSPKKAIPIRANTQEARPTPVKN 981
Qy      2054 -----EPVSVSSPSLTHDKGLPKHLBELDKSHLEGLRPKQGPVYKLGGEA 2100
Db      982 PVKKMKPPWEDDDPEAEVSAPE--PEKKT-----VLAK--KAPAKRDPSPKAAAPVA 1032
Qy      2101 A-----HLPHLEPLESOP-----SSSPLQTAAGVGHQVTLAHOISVI 2143
Db      1033 AKRPPKLPVPPPTPVKAPVKKKKPPWEDDEPSEVSAPE--EKKTPVLAK--KAP 1085
Qy      2144 TQDYRHHPOQLSAPLPAFLYSPGAS--CPVLRLRP--PSDLYLPP-----PD 2189
Db      1086 TKPATKPDSEAAAPVSGPTSKDPLKSLKAPV-EKPKTTTPKODKLKPSAKKPEKAP 1144
Qy      2190 HGAPARSGPSHSGGKRSPEPKNT---SVLGGEDGIEPVSPRGMTPEGHSRAVYPL 2245
Db      1145 PAAKPKMKPVWDDPDDEADFTVPAPSKKPDTEPPADPLGPK-TKDPKLNKAP--A 1200
Qy      2246 YRDEQTEPSMGSKSPGNTSQPAPAFSKLTESNSAMWKSQKQELNKLNTHNNEPEYN 2305
Db      1201 EKPTKPKPKPV-KSEPKPTPEP-----KPAAPKMKPPWEDDP-- 1240
Qy      2306 ISQPTETFFNMPA 2318
Db      1241 --BEPAD-FTMPA 1250

RESULT 28
15116
NF-180 - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: 151116
R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A:Reference number: 151116; MUID:95287814; PMID:7770000
A:Accession: 151116
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1110 <JAC>
A:Cross-references: EMBL:U19361; NID:G632548; PIDN:AAA80106.1; PID:G632549
C:Superfamily: neurofilament triplet H protein

Query Match      3.0%; Score 403; DB 2; Length 1110;
Best Local Similarity 22.7%; Pred. No. 3.3e-07;
Matches 234; Conservative 153; Mismatches 380; Indels 264; Gaps 44;

```

```

Qy      166 RLSEKELLQNMDRVDEITWVEQOISKLKKOQLEBEAA-----KPEPEKPVSPPI-E 220
Db      97 RVNEKQLLDPLN--DRFAGYIE-KVHDLQKKNKLEFETISAVRGQSGAPAGGVSDVYE 153
Qy      221 SKRSLVQIYDENRKA-----EAAHRLBGLGQVEL-----PLYNPSDTRQ 265
Db      154 QEIKELRDVIDIDINGEKTVOIQEHLDEEIQRLREKTDDEVRLNTEALIN-----A 207
Qy      266 YHENIKINQAMRKUL-----ILYKRNH-----ARKW-----294
Db      208 FRKNVDUTSLVRMEMDKRTQSLDEITFLKKNHEEVEDLALQIQSSTVSVERDPAVPE 267
Qy      295 -----KQFCORYDLMEALEKKVERIENNPRRAKESVRE 331
Db      268 ITAALREIRQLBEGSARNIETAEWFKGKSQUTEAEE-----QNNDAIRSAKEITE 321
Qy      332 YTEK--QFPEI-----RKQELQRMOSRVQGGSGLSMAASEHVS-----373
Db      322 HRRKLQRCETELDLAQTESLERQLSEMERHQSVDG---NLQDAQOQLENELRMTKW 377
Qy      374 EIIDGLSE-QENLEKQNR---QLAVIPMLYDADQQRIFKFINMGMLADPMKYVKDRQVM 429
Db      378 EMARHLREYODLNVKKAALDIEIAVRYKL--DGEIRY--SSGLPTPAKPPKAPSAK 432
Qy      430 NMASEQKETEFREKFMQHPKNFGLIAGFLERKTVAGCVLYYLTKKNENYKSLVRSYRR 489
Db      433 PKAKVEKAVYSK-----PEIKYSEPIISNQDITDLDLQEEVMEKAA 478
Qy      490 RGSQOQOQOQOQOQOQOQOQPMFRSSQEEDEKEKEKAEKKEKEKEVENDEKDLKEK 549
Db      479 PVSAEKQDEEEEEEKEEKEEAEAEEDRKRKEGAEEAEAEA-----EEVEKEE 532
Qy      550 TDDTSGEDNDEKAVNAGKRTANSQGRKGRITRSMANVAN--SEEAITPQOQAEI-- 604
Db      533 AEAEVEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 592
Qy      605 -----ASMELNERSMTWEEMETAKGILBHGKRWMSAIAMVSGSKTYSQCNIFYNYKR 659
Db      593 VEEREAFAEEBEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 641
Qy      660 QNLDEILQHKLXMKERKNARRKKKKAPAAASEEAPPVVED-----702
Db      642 EAEEEVTSKAKAQEAEVEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 701
Qy      703 -EEMEAAGVSGNNEEM-----VEEAALHAGNENVPRECGSPAVNNSSDTEIIPSHTE 757
Db      702 AEEDAEAEAEVEEYKESVTKSDAEAEA--EAEBEAKSEBEAEAEAEAEAEAE--EAE 756
Qy      758 AAKDTGONGPAPATLADGPPPPPTPPRTSPAPIETPPASEATGAPTPPPAP-----812
Db      757 AVEETEAATEAEAKKASDDEKPEEY--KESAAFPAP---EAKKAPKAPKAPKAP 809
Qy      813 ---PSPAPPPVVKKEKEE--TAAAPVVEEGEQQPPAAEBLAVDTGAEEFVKSECT 867
Db      810 AKVESPTEPEDEPKAVEKEKGAFAAPK---PKAPPAAKKAKKAVEKEEKEEESP-T 864
Qy      868 EAAEAGAKGDAAATAATAGALKAKKGGSGRATTAASSGAPQSDSSATGSADEVD 927
Db      865 EEEPKKPAKAPAK-APAKPPAPKAAEE-----KPEPKPAQAPAPAAEEED 914
Qy      928 EAEGGDKNRLSPPSLITPTGDPANASPOKPLDKLKORAAAPPIQTVKHEPRE 987
Db      915 EKEDF-----EEEVEVEYKPPDAKPVASKPA-----PAKE 946
Qy      988 DAAPTGA--PPAD---PPONTQPESDA--POQPGSP-----RGKSRSPAP-PAD 1031
Db      947 EEDEPKAKQPKKPKRPPAPKEPEDEKAPKAKKHSFVEERKPIKEIAPAKAAPAKAD 1006
Qy      1032 KEAPAAEAQKL 1042
Db      1007 KEPPAAEPKTI 1017

```

```

Db      1569 GAGMLHITENGPTEV-DY-----SPSDI 1590
Qy      1732 IDLSQVHLPLVLPPTGTPTATMDRLAFLPTAPQPPSSSRSSPLPGGPTHLTKPTTT 1791
Db      1591 QDS-----LSHKLPTEPEPTQNDLSELISQVSAPESSAHPSQIASPLQDITL 1646
Qy      1792 SSSREDRD-----REDRDRERE-----KSILTSTTVEHAPIMRPG-TEOSSGSSGS 1840
Db      1647 SDVPPDMSLYASLASEKQSLSEKLSPEKSDISPLTPRESSPTTSPGFSDSITGAKES 1706
Qy      1841 SGGGGSSSRPASHASHAHSPISTQDALQORPSVLHNTGMKGIIT-AVED----- 1892
Db      1707 TAAVQTSPPIDAAAEPEYGFSSMLFTMQ-----HHLASRLDITTSVEKNDGKT 1760
Qy      1893 -----SKPTVLKSTSTSPVPAATFPPTATCPGLGTLGVYPTLMEVLLPKAPR 1944
Db      1761 PGDFNAYQKP-----ESTTESPDEEDYDSHEKTIQAHVGGYY-----YEK 1804
Qy      1945 VAPPERPADTCHAF-----LAKPARSL-----BPASSPSKG-----SEPRPLVP 1987
Db      1805 TERTIAPCSGSIYETIEKTKTPEDGYSCEITETKTRPEEGYSYEISEKTRTPE 1864
Qy      1988 VSGHATIAKTPAKNLAPH-----HASPDPAPPASADPHREKTSKP---F 2031
Db      1865 VSGYVTEKTERSRRLDDISNGYDTEDEGHTLGD-----CSYETETKITSFPESESY 1919
Qy      2032 SIQELERISLGYHSSSYSPGCVFVSFVS-----LTHDKGLPKH----- 2073
Db      1920 SYETTTKTRSPDTSACYEEMKTKTPQASTYETSRYCSEPKESSEARQVDLC 1979
Qy      2074 -LEELDKSHLEGELRPK--QGPVK- LGSEALHPLRLPESSPSSPLQTAQVKGH 2129
Db      1980 LVSSCFERKHEKTELSPFIPNPLFEMFAGEPTESERPLTQSGAPPP--SGCKQQR 2036
Qy      2130 QRVVTLAQHISEVITQDTRHHPQOLSAFLPAPLYSPPGASCVP-LDLRRP---PSDLY 2184
Db      2037 QCCEPTPTSVSE-----SAPSGTDSVDPETRECPSTADANLSDSESETIPDKT 2088
Qy      2185 L-----PPPDHGAHPANGSHSBECKRSPENKTSVLGCGEDGIEPVSPREGMT----- 2232
Db      2089 VTYKMDPP--APMO-----DRSPSPRHDPV-----SAWDEBALAIEQNLG 2128
Qy      2233 -----EPGHSRAVYPLLYPDGEQTEPSRMGSKSPGNTSOPAPFSLTES 2278
Db      2129 KALKKDLKERAKTYPGTITKSSPVKKGDK-----SKPSAASPKPGALKSSDK 2179
Qy      2279 NSAMVSKQOE-INNKLT 2296
Db      2180 VSRVAPSKKESVEKAMKT 2198

```

RESULT 27

```

T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leimach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U6675; NID:g116613; PID:g116621; PIDN:AA52641.1; GSPDB:GNOC
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP.F35A5.1
A:Map position: X
A:Insertions: 1272/2

```

```

Query Match      3.1%; Score 404; DB 2; Length 1274;
Best Local Similarity 20.1%; Pred. No. 3,6e-07;
Matches 353; Conservative 167; Mismatches 601; Indels 632; Gaps 87;

Qy      682 KKKKAPAAASEAAPVVEDEEMASGVSGNEEVEAEALHASGNEVREGCGPAT 741
Db      14 KKKKP-----WESVDEEEMEDV-----ETPA-----PSKLEKKPSTL 47
Qy      742 VNNSDTESTPSRHTTEAKTQNGNRPRTATLGADGPPPPPTPRRTSRAPTE-PPPAS 800
Db      48 KKKDAPKVPSP-----DA--PSVPINPVKKWAPWEDDEPME 86
Qy      801 EATGAPTP-----PAPPSAPPPVVPKEKEEETAAPVREGCEOKPAA- 848
Db      87 EAPAAVPKAVQDPSKPKVPKAPKPRDASPCKIAAKKEPTTLAVPTPVKPVKKAP 146
Qy      849 -BEIADVTKAESEPVKSECTEAESEBPAGKDAEAAEATAGALKAKEKGGSGR----- 902
Db      147 WEDDEVV-----EDVKDAPTPAKKTPVLKKKPPAAAKPRD--PSPKAAPSEKHPIV 200
Qy      903 -----ATTASSGAPQDSSSATCSADEVDEAGGDKN--RLISPRL-LTPGDPAN 954
Db      201 PPTPIKPKAKKMPWEDDEVPT--EIKEBPATRKVPALKKEPSTSVKPVSD- 254
Qy      955 ASPQKPLDLKQKORAAAIPIQV---TKVHEPPREDAPTPAPAPPPONLQPSDA 1011
Db      255 -SPKKVPVK---KEEVVPTPIKPTKMKKPPWEDTEVEYKPP-----VPEKKA 303
Qy      1012 P--QOPGSSPRGSRSPADPKAPFAEAKLPGDPWCWTSGLPFPVPPREVIKASPHA 1069
Db      304 PVLKKDPAAPAAKARDSPS-----KAAPKK 329
Qy      1070 PDSASVYAPRGRPLPLGLHDTARPVLPRLPTISNP-----ISSAKHPSVL 1118
Db      330 VEPSS-----PVV-PTPVKNPVKKYKPPWEDDEPAEEVKKSPAP 369
Qy      1119 ERQIGATISQMSVOLHVPSEHAKAPVGYTMGLPLPMDPKKLAPFSGVQEOQLSPRQA 1178
Db      370 EKTTPVLAK-----KEPSSSTTSSDP-SPKKAAP-AVAPRSSPK-KA 411
Qy      1179 GPESLGVPTAQSASVLRGTALGASVPGSITKGI PSTRVPSDAITVYRGSITHTGPADV 1238
Db      412 TPLQA--DPAQD-----VPTPVKNPVKKYKPPW----- 440
Qy      1239 YKGTIRIIGEDPSRLDRGRSDSLPKGHYTGKKGHVLSYSGMSVTCSEKDGSSS 1298
Db      441 -----VDDEDP--VEEVKQEPAP-----AKTTPVLK-----RKEPAKOT 473
Qy      1299 GPPHETAPRKRTYDMMEGRVGRAISSASIEGLMGRAIPRRHSPHILKEQHTRGSITQG 1358
Db      474 AKPATSKTP-----TPKKDPPVRRSS----- 497
Qy      1359 IPRSYEAOEDYLRRKALKREGTPPPP-----PSRDL--TEAYKTQALGPLLKP 1409
Db      498 -PKK-VAAPKDSAQAPATPVKNPVKKRPPWEDDETPADVDSKPTAKKTPSLAKKDPAP 555
Qy      1410 AHGELVATVEAGRSIHETREELRHTPBLPLAPRLKESITQGTPLKTDGTASTGSK 1469
Db      556 AKESLK-----PKADTK--APAKRPDPSPKV-----APTAPKK 587
Qy      1470 KHVRSILSGRGRTPPVHPLDVMAADALERACYESLSKSRGTASSSGSIRAGAPVI 1529
Db      588 KTFVLA-----KKEPAGP--ADSKTYS----- 607
Qy      1530 VPELGRKROSLTYEDHGAPFAGHLPKRGSPVTKREPTPLQEGSLSSSKASQDRKLTSTP 1589
Db      608 -PEKSKPRD-----PSPKAAVPAKVPKTEVAPAAVKK-----P 640
Qy      1590 REIAKSPHSTVPRHHNHPISTPYHLLRGVSGVDLYKSHPLADPISIRGI-----PLD 1644
Db      641 EPISK--PKDTAPK-KAEPNSV-----VP-----PTPVKNPVKKKPPW 678
Qy      1645 AAAAYVLRHLAENPTVPHLYPPYLIRGVYDPTALENRQITINDYITSQOMHNTATAMA 1704

```

RESULT 26
 A56577
 microtubule-associated protein MAP 1B - rat (fragment)
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997
 C:Accession: A56577
 C:Author: W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.
 Eur. J. Cell Biol. 57, 66-74, 1992
 A:Title: Identification of two distinct microtubule binding domains on recombinant rat M
 A:Reference number: A56577; MUID:92347374; PMID:1639092
 A:Accession: A56577
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2364 <ZAU>
 A:Cross-references: GB:X6050
 A:Experimental source: brain
 A>Note: nucleotide sequence not given; conceptual translation not complete
 C:Superfamily: microtubule-associated protein MAP1B

Query Match 3.1%; Score 406.5; DB 2; Length 2364;
 Best Local Similarity 19.3%; Pred. No. 5.7e-07;
 Matches 489; Conservative 326; Mismatches 995; Indels 729; Gaps 107;

QY 68 ILISEQPNERSQEL-----HLRSHSYLPEL-GKSEM---- 100
 DB 79 LLSSTHPANKASTLFCPEBGDMKNSNLDNRHNLDPINIKLNSASILEMEGLSEFTTEYL 138
 QY 101 -EFIESKRPRLLELDP-----LLRP-----SPLLATGQ-----PAGESD-- 134
 DB 139 SESSEVPSRP-FDILEPPTSGGFLSKSCCYIFPGRGDSALFANGNMMLINGSSEKKS 197
 QY 135 -----LTK--DRSLTG-----KLEPVSPSPPHTD-----PE 159
 DB 198 CFWKILIRHLDVDSLILTHIGDNLPGINSMLQRKIAELSEBSQSTNSDMMKMLISPD 257
 QY 160 LEL-----VPRRLSKELLQMDRVREITMVEQOISLKKKQOQLEBEAAKRP-----E 211
 DB 258 LGVVELANVPENLKNBPNIKMRSTEEACFTLLQYLNKLSMKRPELFRSVGNAIEPVILFQ 317
 QY 212 K-----PVSPPTISKHSLVOITYDENRKKAEAAHRIEGLGPOVELPLYNQPSD 262
 DB 318 KMGVGLKMYULNPNYKSSKEMQYPMQWTGKDAE-----LILNGQEVDP----- 366
 QY 263 TRQYHENTKINQAMRKULLYFKRRNHAKQKQKQFCQRYDQ--LMEALEKKVERIENNP 320
 DB 367 -----ISYLAVSLSLIWMHPANPAEKIRVLFPGNSTQYNILEGLE-KLKHLDPLK 416
 QY 321 RRRAKESKRYREYEQFPIRKORLOEMOSRVQORSGLMSAARSEHVESEIIDGLS 380
 DB 417 QPLATQKULTG--QVSTPVPV-KQVTLKQKQADRESLKPATKPLSSKVRKESKEAPKAT 473
 QY 381 EOENLEKQWRQLAVIPMLYDADQORIKFINNGGLMADPMKYKDR--QVMNMSEQEK 438
 DB 474 KASQYK-----TPKVESKEKVIK-----KQKPGVESKPEVTEKE 510
 QY 439 TFRKFMQHPKNFGLIASFLERKTVACVLYYLTKQENYKSLVRSYRRRGKSQOQOQ 498
 DB 511 V-PSKEQSP-----VKAEVAEKAATE-----SKPVTKDKVYVKEIKTTPREKKEK 557
 QY 499 QOQOQOQOQOQPMRRSSQEEKDEKKEKEAKKEBEKEPVENDXKDLKEKTDUTDSGN 558
 DB 558 PKKEVAKEDKTPPL-----KQDEKPKKEAKKEIKKEIKKEKELKKEVVKETPLKD- 610
 QY 559 DEKEAVASGRKTANSQGRKRGRITRSMANSEBALTPOOSAEIASMELNESSRWTEE 618
 DB 611 -----AKKEVKKDEKKEVKEEK-EPKKEIKKISDKIKSTPILST 650
 QY 619 EMETAKGGLLEHGRNMSALARMVGSKTVSQCKNFFNYKKRONLDEIIOQHLKMEKERN 678
 DB 651 KKPAAALK-----PKVAK-----KEEPTKKEPIAAGTKDKGKVK 684

QY 679 ARRKKKKAPAAASEAAPPVEDEMEASGV--SGNEEMVEBEALHAASGNEVPRGEC 736
 DB 685 VIKKEGTTERRAAAVGTAAV-----AAAGVAASGPAKL-----EAKSLM 727
 QY 737 SGPAVNNSSDTSIESIPSPHTEAKDYGQNGPKPPATYGADGPPPPPTPPARTSRADIEP 796
 DB 728 SSPEDL--TKDFELKAEKIDVAKDI-----KPOLIELID-----EELKKE 766
 QY 797 TPBAEALGAPTPPAPSPAPSPAPPPVVPKKEBEETAAAPVE-----EGEROKPAAEEL 851
 DB 767 TEPEEAV-----VIKETEVSIGSASPEDEGITTEGESECCQTEBEL 809
 QY 852 AVDTGAEEPYKSECTEAE-----EGPAKGDAAEA--EATAEGALAEKKEGGSGRATT 905
 DB 810 -----EPVEKGVNDIEKFEDEGAGFESSSAGYEKAELEAEAEPEDEDEDNVSG 861
 QY 906 AKSSGAPQDSSSATCSGADEYDEAGDKNRLSPRSLTPTTGDPAAANSPOQPLDLKQ 965
 DB 862 SASKHSPTDEBEIKAAADV-----HIKERESEV--ASGDDBAEEDMDALEKGE 909
 QY 966 LKQRAAALPPIQYTKNHEPPREDAPTK-----PAPAPPPQNLQESDAPQ 1013
 DB 910 AEQ--SEBEGEEDKADAREEDHEPDKTEADYVMAVVDKAAEAGVTEDQYDPLGTPAK 968
 QY 1014 QPG--SSPRGKSRSPAP-----ADKEAFAEAQKLPDPPCWTSGLPFPVPREVI 1063
 DB 969 QPGVQSP--SRPAASSIHDETLPGSESEKATADENRBDQPEEFTATSGYTGSTIEI 1024
 QY 1064 KASPHADPSAFYAPGHPILGLHDTARVLP--RPTTISNPPLISSAHPVYLE 1119
 DB 1025 -----SSEPTPMDEMSTPRVMTDETNNETBESPOGFVNITTEESL- 1067
 QY 1120 RQIGALISQMGVQAHVPSHAKAPVGVYTMGLPMDPKKALFSGVQKQOLSPRG--Q 1177
 DB 1068 -----YSQEYSKPVVAASFGSLDSGSKTDATDGRDYNASASTISPPSSMEDEKFKSALRD 1122
 QY 1178 AGPPESLGVPPAQAASV-----LRGTALGSVPGGSLTK-----GIPSTRVSDS 1221
 DB 1123 AYRBEETDVTKGALDLIDVDSERLSPAKSPSLSPSPSPPIEKTPLGERSVNFSLTPNEI 1182
 QY 1222 AITRGSITHTPADVLKGTITRIIGEDSPSLDRGREDSLPKGHVITYEKKGHVLSYE 1281
 DB 1183 KASAEGEAT-----AVSPGVTAQVVEHCASPEKTLLEVVSFQSV--TGSAGHTPYQ 1235
 QY 1282 GGMSTVQCSKEDGSSSGPPHET-----AAPRTYDMEGVRGAIISSAS-----IEG 1329
 DB 1236 -----SPDKSSHLPTVTENAOAPV--SEFTFAKBNERSISIPMDEVPDSBS 1286
 QY 1330 LMGRAIPPERHSPHNLKEQHHIRGSITOGIPRSYVEAOEDYLRRBAKLLKREGTTP-- 1385
 DB 1287 PIEKVLSPLRPP-----LIGS-----ESAVEDPLSADDKALGRSESPFEGK 1329
 QY 1386 -----PPPSRDLT-EAYKTOALGPLKAPNAHEGLVATVKAAGRSITHEIPREELHT 1436
 DB 1330 NGKQFSDKESPVSLTSLDYOD-----KQEBKRAAGPIPKEDFSPKKAADAIMSSQ 1383
 QY 1437 PELPLAPRPL-KEGSIITQGTPLKYDTGASTGSKKHUVRSLIGSGRTFPVHPPLDVAD 1495
 DB 1384 SALADEKRLGDDGPTQ-----VDVSQGSFKEDTKMSISGTVSDKATPVDEGA- 1435
 QY 1496 ARALEACYESLSKRPGTASSGGSILARGAPVILPGLKRGROSLTYEDHAGAPACHLP 1555
 DB 1446 -----EDTYSMEGVAVSTASVATSS--FPPEPTDVSGLHABVGSFSTHVD 1483
 QY 1556 RGSPLYTKEPTRLQESGLSSKASQDRKLTSTPREI--AKSPHSTVPEHHHPPIPY 1611
 DB 1484 DSLASVAVQUTPTTFQETMSMSKCECPRMWISIPDPFSPKTKAS--RTVQDHRSE- 1537
 QY 1612 EHLARGVGVDLYNSHPLADPTPSIRGICFLDAAAAYLLPRHLAPNTYPLHLYPYLIR 1671
 DB 1538 -----QSMKSIIEFGQSPESHLMDS-----ROSFDHFTV----- 1566
 QY 1672 GVPDTAALENROTIIINDYITSQOMHNNTATAMAQRADMLRGLSPRESSIALNYAAGPGI 1731

A:Accession: 155254
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388
R:Bonsstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control
A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
J. Biol. Chem. 260, 2315-2320, 1985
A:Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter
A:Reference number: 155237; MUID:85130970; PMID:2857713
A:Accession: 155237
A:Molecule type: DNA
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Residues: 1-34 <CH2>
A:Cross-references: GB:M0627; NID:G180383; PIDN:AAA51992.1; PID:G553226
R:Wirtz, M.K.; Keene, D.R.; Horst, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A:Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
rome, type VII.
A:Reference number: A35233; MUID:90202908; PMID:2318855
A:Accession: A35233
A:Molecule type: Protein
A:Residues: 33-52 <MIR>
A:Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R:Well, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A:Title: A base substitution in the exon of a collagen gene causes alternative splicing
A:Reference number: S09400; MUID:89356643; PMID:2767050
A:Accession: S09400
A:Molecule type: mRNA
A:Residues: 156-183 <MBI>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A:Reference number: A30567; MUID:71038625; PMID:5529814
A:Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
A:Accession: B90567
A:Molecule type: protein
A:Residues: 162-196, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
A:Experimental source: skin
A:Note: evidence for 170-allysine
R:Batge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F.
Eur. J. Biochem. 192, 153-159, 1990
A:Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A:Reference number: S11372; MUID:90382436; PMID:2169412
A:Accession: S11372
A:Molecule type: Protein
A:Residues: 175-187, 274-287, 'P', 289 <BAE>
A:Note: sequence of collagen alpha 1(S) I isolated from bone after pepsin digestion
R:Deak, S.B.; Scholtz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A:Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
operative melting of intact type I collagen.
A:Reference number: 155342; MUID:92042092; PMID:1718984
A:Accession: 155342
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 258-266, 1347-1357 <DEA>
A:Cross-references: GB:S67495; NID:G239007; PIDN:AA20350.1; PID:G239008
A:Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A:Reference number: A92069; MUID:71001508; PMID:4319110
A:Accession: A92069
A:Molecule type: protein

A:Residues: 263-268 <MOR>
A:Experimental source: skin
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R:Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A:Title: Segmental amplification of the entire helical and telopeptide regions of the CD
A:Reference number: S15989; MUID:90326017; PMID:2374517
A:Accession: S15989
A:Molecule type: mRNA
A:Residues: 281-302, 402-420, 823-843, 925-944, 1026-1045, 1143-1162 <LAB>
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Preclorius, P.J.; de Vries, W.N.
Connect. Tissue Res. 29, 1-11, 1993
A:Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A:Reference number: 152905; MUID:93339042; PMID:8339541
A:Accession: 152905
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 342-352, 'C', 354-359 <W12>
A:Cross-references: GB:S64717; NID:G408195; PIDN:AA27677.1; PID:G408196
A:Note: mutant sequence from patient with osteogenesis imperfecta
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Bikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A:Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha1
A:Reference number: A50476; MUID:84080385; PMID:6689127
A:Accession: A50476
A:Molecule type: mRNA
A:Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
A:Cross-references: GB:K01228; NID:G180391; PIDN:AAA51995.1; PID:G180392
A:Note: sequence partially completed for missing nucleotides by A29439
R:Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A:Title: Multicodon deletion in an osteogenesis imperfecta variant with increased type II
A:Reference number: A22161; MUID:85104934; PMID:2981843
A:Accession: A22161
A:Molecule type: DNA
A:Residues: 472-594, 'R', 596-607 <CH3>
A:Cross-references: GB:K03178; GB:K03179; NID:G179612; NID:G179613; PIDN:AAA51847.1; PID
A:Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R:Forlino, A.; Zolozzi, F.; Valli, M.; Pignatelli, P.F.; Cetia, G.; Brunelli, P.C.; Mottes
Hum. Mol. Genet. 3, 2201-2206, 1994
A:Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
A:Reference number: 154365; MUID:95187161; PMID:7881420
A:Accession: 154365
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 746-766, 'S', 768-781 <FOR>
A:Cross-references: GB:I47667; NID:G1009093; PIDN:AA59576.1; PID:G1009094
R:Chesler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A:Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A:Reference number: A47426; MUID:93352646; PMID:8349697
A:Accession: A47426
A:Molecule type: mRNA
A:Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
A:Cross-references: GB:S64596; NID:G407589; PIDN:AA27856.1; PID:G407590
A:Note: sequence extracted from NCBI Backbone (NCBI:136444, NCBI:P136445)
A:Accession: B47426
A:Molecule type: mRNA
A:Residues: 1179-1464 <CH4>
A:Experimental source: normal dermal fibroblast culture
A:Accession: C47426
A:Molecule type: mRNA
A:Residues: 1179-1276, 'H', 1278-1464 <CH5>


```

QY 317 ENNRRRAKESKVEYVEYKQPEIRKQRELOERMOSRVGQSGSLSMGAARSEHEVSII 376
DQ 94 QTSDDHSDTSS-----PEVRQSHSESPSLQSKQSTPKG---GRSSSSPEVE-- 139
QY 377 DGLSEONLEKOMEQLAVIPML-----YDADQRIKFINMGMLADP--MKYKD 425
DQ 140 --LASRPRIRODRGEFSA--SPMLKSGMSPQSRFQSDSSYPVTDNSNLLQSLRETFAS 196
QY 426 RQVNNMMSSEQKEKF-----REKFMQHKNFGLASFLERKTVAECVLYLTKNENYK 480
DQ 197 KEXMAL--PPQEDATAPEPQDKKSPFP-----VQDRPESLV-----FK 235
QY 481 SLVRSYRRRGKQQQQQQQQQQQQQQQMPRSPQSEKDEKEKEKAEKEEKEEVEN 540
DQ 236 DTLATPRERSGA-----GSPETKEQNSALPTSSQ---DELMVEYKEEPEAGQLS 286
QY 541 DKEDLLKEXTDUT--SGEDNDEKAVASKGRKTANSQGRKGRITRSMANEANSEAITP 598
DQ 287 HLSSELKEMSTNSPESPEVERPAV-----SLTL 316
QY 599 QQSLELAMELNE-----SSRW-----TEEMETAKKGLLEGRMWSAIAKRVGSKTVSQC 649
DQ 317 DQSQSQASLEAVEVSPMASSWGPHFSEPEKELNSPLREN--SRGSPLEFRNGPLGTE 374
QY 650 KNFYFNKKRONLD--EILQHKLMKEKERNARKKKKAPAAASEEAAFPVVEDEMEAS 708
DQ 375 MNTFSSSEVKDLNGPLNQ--LETDPSLDKMEQSTSSGHSSE-----LSDVAEKA 426
QY 709 GVSQNEEMVEEALHASGNEVPRGECSPATVNSSDTESIPSPHTEAKDYG--QNGP 767
DQ 427 GMSNQSISSPVLDAVPTPRERSSSSASPEMDGLPRT---PSRRSRSSSGPLDGS 483
QY 768 KPAPTLGADGPPPPPTPRRTSRAPLEPTASEATGAP--TPPPAPSPSPAP---PPV 822
DQ 484 GTPRHSLSGSSPEMKDIPRTPSRGRSECDSPPEPKALPQTPRPRSRSPPSELNKKLT 543
QY 823 PKEEK-----EEETAAPVEEGEBOQKPAEELAVDTGKAEPPVSECTEEAEKPA 875
DQ 544 PQRSSSESSVDQKTAKRTL--QQRSSGSSQELDKPSASQOE-----RSSSDSP- 595
QY 876 KGQDAEAAEATGALKAKEKGGSGRATTAKSSGAPDDSDSATCSADVEDEAGDKN 935
DQ 596 --DSKAKTRT--PLRQSRSG-----SSPEVD-----SKS 621
QY 936 RLSPRSRLTPTDPPRANASPOKPLDKQKORAAIPIQVTKHHPREDAPTKPA 995
DQ 622 R-LSPRS-----RSGSSP---EYKD--KPRAA-----PRAQSGSDSP 655
QY 996 PPAPPPQNLQPESDAPQQSSPRGSRSPAPADKEAFAPAEAOQLPGDPPCMTGSLP 1055
DQ 656 PKAPAP-----RALPRRSKSSSSKGRGSPBG-----SSSTRESSBHP----- 695
QY 1056 PVPREVIKASPAHPDPSAFAPPGHPLPLGLHDTARPVLPRTTISNPPILISSAKHP 1115
DQ 696 --KSRTARSGRSSEPEKTSRTP-----PRRRSSKSSSELTRKAR-- 734
QY 1116 SVLERQIGALISQMSVOLHVPYSEHAKAPVCGVPMGLPMD---PKLAPFSGVQEO 1171
DQ 735 --LSRRRSASSSPDETSRTP--PHRRSP-----SVSSPEPAEKSSRRRRSASSPRTKT 787
QY 1172 LSPPGQGPPESLGVPFAOAS---VLRGTAIGVPGSGITKGIPTRVPSDSAITVR-- 1226
DQ 788 TSTRGRSPSPKPKQLQSRSSRSREKRTTTRRRRSGSSQSTSRRRQSRSSRYTRRR 847
QY 1227 -GSITTG--TPADVLYKGTITRIIGEDSPSLDGRSDSLPKGHVYEGKKGHVLSYEGM 1284
DQ 848 GSGSVHSRSPA-----RQESSRTRRRRRRSRTP----- 878
QY 1285 SVTQCSKEDGSSSGPHEHTRAP--KRTYDMMEGVGAISASIGLMGRAIPPERISPH 1343
DQ 879 ---TSRRKRSRSTSP-----APWKRSRSRASAPATHR--RSRSTPLISRRRSRSTSPV 927

```

```

QY 1344 HKEQHHRIGSITGIPRSYVEAOEDYLRE-----AKLKEGTPP 1385
DQ 928 S-RRRSRSTSVTR--RRSRASPVSRKRSRKTPTVTRRRSRSTPTTRRRSRSTPP 984
QY 1386 -----PPPSR-----DLTEAYKTQALGPLKPRANEGVATVYKAGRS--IHE 1427
DQ 985 VTRRRSRRTPTVTRRRSRSTSPITRRRSRSTSPVTRRRSRSTPTVTRRRSRSTSP 1044
QY 1428 IPREEL--HNP-----ELPLAPRLKEGSIQSTQPLKY--DTGASTGSKHDVRS 1476
DQ 1045 VTRRRSRRTPTAIRRRSRSTPTPLPKRSR--SPLAIRRRSRRTPTARGKSL 1100
QY 1477 IGSPPRTFPVHPVADADALERACYESIKSRPTASSGSGSIRGAPVYVE----- 1532
DQ 1101 TRSP-----PAIRRSAGSSSDRGRSA-----TPPATRNHG--SRTPVALNSSRMS 1147
QY 1533 -LGRQSPPLYEDHGAPFAGHL--PRGSPVTRREPTPLQOE--GSLSSKASQ--DRKLT 1586
DQ 1148 CFSRPSMSPPLDRCRSP--GMLEPLGS--SRTPMSVULOAGGSMMDGPGRIPDHQR 1202
QY 1587 STPREIAXS-----PHSTVPEHHPPI-----SPYEH 1614
DQ 1203 SVPENHAQSRIALALTAISLGTAARPPSMGAGLAAMSOVPA--PVPLMSLRTAPANT 1260
QY 1615 LRGVGVDLVRSHTPLAFDPTSPRGIPL-----DAAAYVL--PR-HLAPN----- 1658
DQ 1261 ASRTPAASAAAMNLASRTP--AIPTVANLADSRTPAAMNLASPTTAVAPSANLADP 1319
QY 1659 --PTYPHL-----RYPDTALENQTIINDYITISQOQHNTATAMA 1704
DQ 1320 RTPAPAVNLAGARTPALAALSLTSGTPEPTAA-----NYPSSR-----TPQA 1364
QY 1705 QRAMLGLSPRESLANTAA-----GRGIIDLSQVPHLPVLPVPPPTPATMDR 1757
DQ 1365 PASANLVGPSAAHATVANIAGSRTAALAPASLTISARMAPALSGANLTSPPVLSAYR 1424
QY 1758 LAVLPAPOPFSSHSSSPLSPGPTHLTKPTTSSSERDRDR--RDRDR-- 1808
DQ 1425 VSG--RTSPPLDRAKSTTPSABQSMSTSRAPSPSRKQAPSGCLLPAPAOQPSPV 1443
QY 1809 -----EREKSLTSTTVEHAPIMRPGTEQ--SSG-----SSGSGGGGSSSRPASHAH 1858
DQ 1484 PSAPSDQRCILIAQTIV-----AGQSLSGCAVATTTSSADHNGMLSVAP--GV 1533
QY 1859 QHSPISRTQDLDQORSVLHNTGKGIITAVESKRTVLRSTSTSPVPAALFFPATH 1918
DQ 1534 PHSDVGEPPASTGAQPSAL-----AALQPAKERSSSSSSSSSSSSSSSSSS 1584
QY 1919 CPLGGLTDG-----VPEPLM-----EPVLLPKCAPVVARPERP-----RADTGHA 1959
DQ 1585 SSSSGSSSSSEGSILPVQPEVALKVPSTPAKPAKAVREGRPPEPTPAKRRSSSSSS 1644
QY 1960 LAKPPARSGLEPASPSKSGSEPRPLVPVSGHATIAATPAK--NLAPHASDPPAP-- 2014
DQ 1645 SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1704
QY 2015 ----PASADPHREKTSKSPSIQLEL--RSLG-----HSSSVSPREVEVSP 2058
DQ 1705 RSLLPVSPSPRH-----SLPHVARGTFLKRTSGVFPPLAKSFGLQDAP 1752

```

RESULT 24
 S15053
 hypothetical protein YCR033w - yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein YCR592
 C/Species: Saccharomyces cerevisiae
 C/Date: 07-May-1993 #sequence revision 07-May-1993 #text_change 19-Apr-2002
 C/Accession: S15053; S19445; S60918
 R/Jia, Y.; Slonimski, P.P.; Herbert, C.J.
 Yeast 7, 413-424, 1991
 A/Title: The complete sequence of the unit YCR59, situated between CRY1 and MAT, reveals
 A/Reference number: S15052; MUID:91335897; PMID:1872032
 A/Accession: S15053

```

Qy 811 APPSPAPPVVPEEKEEETAAAPVEEGEE-----OKPAAEELAVDTGKAEEPVKS 864
Db 2523 ASPAPSAESPFI--QAYKOESQEHNSLEQOQSRSSVSHEDSPAAQYHNDSDHSDSPVS 2580
Qy 865 E---CTEAEGBGPAKGDAAEAETAAGALKAKEKKG--GSGRAATKASGAPODSDSAT 920
Db 2581 DRAFLLEQOQHQPESGEESD--GEGFGSKVLGFAKKAGVAGVAAVLAALAAVAKAAAYD 2639
Qy 921 CSADEVDEAGGDKRLLSPRLSLTTPTGDPANASQKPLD--LKOLKQRAAALPIQ- 977
Db 2640 ALKKDDDEDENDPEOEKLLPK-----SPERQVAVNVPESDEISIELEYTSPSPTEK 2693
Qy 978 -----YTKVHEPREDA-----APKRAPAP----- 999
Db 2694 SESOCYETVATTTVTREYLDLDDPOSVTRSRSPSEHDISEQYAPESPVEQDPYVEKTTTV 2753
Qy 1000 -----PPONLQ-----PESDAPQOPGSSPRG-----KSR 1024
Db 2754 IROVHDEPPOEIEQOTIPEEVTVAREVESPEDEPQOHIETKTTTITKEVAVPEED 2813
Qy 1025 SPAPPADKEAFABAQULPGDPCWTGSLPPVPPEVYKA-----SPAPDPSAFSY 1077
Db 2814 VOISPVHSETSVSEKQ-LPAD-----EQLDEPTESTTATVTRERYEPEEVPAPSGSE 2867
Qy 1078 APFGHPLPLGLHNDARPLPRPTIS-----NPPPLISSAGHPSVLEROI 1122
Db 2868 ADESHAPKMETTTTTTTRREYEVSEDEDEHQOSQVORDESPAPSEDSVGH-VIEK- 2923
Qy 1123 GALSQGSVOLHVPVSEHAKAP-----VGPVTGCLPLMPD-----K 1159
Db 2924 -TTTTVTREYEDSHSPVPSDDVAGFVKTTTTTTHHEHPEEDPDSDEHVSE 2982
Qy 1160 KLAFGSGVQOQLSPRG-----QAGPESLGYPTQOBSYL 1195
Db 2983 RYASGSPVPSSEDSREIETTTTTTRHEHELEDDEHVESQESYASGSPVSEKSV 3042
Qy 1196 RGTALGSPGGSIRK-----GIP-----STRVSDSARITRGITGTPADV 1237
Db 3043 R-VIETTTNTVTRHEHEHEDLPTIVETSHDPPAASVSEEDV-----HQIOTTT 3093
Qy 1238 LYKQITR--IIGDSPSRDLRGREDLPGKHVIEGKKGHVLSYEGMSV-----TOC 1289
Db 3094 TTTTTRREHVPE-----EIDSGMDELEK-----YSESPVPSSEDSRVIETTTTTTV 3145
Qy 1290 SKEDGRSSGCPHETTAARKTYDMGEGVGAISASIEGLMGRALPPRHSPHHLKEQH 1349
Db 3146 SREHPEED--DHSHVVESQCY-----SAS-----GSPVSEK-----SVR 3180
Qy 1350 HIRGSIQGISPVVEAOEDYLAREAKLKRGGTPPPPSRDLTGAQLGRLKLP 1409
Db 3181 VIETTTTTTRHEHPTBEDTIPESRTSHDDGTTDQHVPSQS----- 3223
Qy 1410 AHGELVATVKEAGRSIHIEPREELRH-TPELPLAPRLKEGSIQTGTPLYKDTGASTTGS 1468
Db 3224 -----PVPEEDDHVTHEQIINDDPIDE-QIYSHSHYKSSSIPSPQD 3265
Qy 1469 KKQDVRSILGSP--GRTPPPPHPLDWMADA-----RALERAQYSESL-----K 1509
Db 3266 STHVIEETVMTSTPITSEKYDPEVEKDVVASADEIDSTAYQYKSESPPVQTSKSLIAKQ 3325
Qy 1510 SRPCTAS-----SSGSIARG--APVIVELG-----KPRQSPLYTED 1545
Db 3326 QESDESDGSEFGSKVLGFAKKAQVAGVVAALPALLAAVAKAAAYDALKODDEDEQED 3385
Qy 1546 HGAPFAGHLPRGSPVTKMEPT--PRLQE--GSLSSSKASQDRKLTST-----PREI 1592
Db 3386 KEPLIGGHHKQODDISODEPSAVDSETOEPGATFPEPEEDKXVITDSADSSVQDEPKIV 3445
Qy 1593 AKSHSTVPEHHPHPISEYHLLRGVGVDLYRSHIPLAFPTSLPRGIPLDAAAYLPL 1652
Db 3446 --FVDSITPEHNSDREBFISIVSG----- 3470

```

```

Qy 1653 RHLPNFTYPLHVPYLIRG--YPTDALENRQTIINDYITTSOOMHNTATAMA-QRADM 1709
Db 3471 -----PIVSESTDVAQTSABEPR-----ISSPVHSDADGASFKRPES 3508
Qy 1710 LRGLSPRESSLAINVAAQPGGIIIDLCOVHLPVLPPTPGTPTANDRLAYLPAAPPS 1769
Db 3509 VTGDEKNA-----IPETSETDAPV-----IDSEEEFN 3538
Qy 1770 SRHSSPLSPGGPHTLTKPTTSSSERDRDR-----RDRDRREKSLTST 1819
Db 3539 NRDBORISSP-----AHSDDEENDAEVIDSEFTRHSQONNEEDSIV----- 3582
Qy 1820 TVEHAPIWRPGTEQSSGSGSGGSGSSSRPASHS-----HANQSPISPR--TQDA 1870
Db 3583 -----ESGEYISSGHG--SPRPEDSTTTVAVNHHPALPEPEVEEE 3625
Qy 1871 LQORPSVIAHTGK-----GIITVAPEKPLVLA NSTSSPVRAATPPATHCPLCG 1923
Db 3626 LEOGRSIISEEEKYSSPLPPTSVTVHEHVPABHKKYRTTSP--IIVTVSSSH----- 3678
Qy 1924 TLD-GVYPTLMEPVLPKEAPVARPERPRADG-----HAPLAKP-PAR 1966
Db 3679 -LDNQPFPVVESEETRASPL--LGPFRPSPSGSLPREEDSHVIESHHTSSVPSPS 3735
Qy 1967 SGL-----EPASSPSKSEPRPLVPVSGHATTAATPAKULAPH-HASPD-- 2010
Db 3736 DSVGHVIEKTTTTVTREREPEDESHSPVSEDDV--HGCVKTTTTTTHHEHPEEDH 3793
Qy 2011 -----PPAPASADPHREKTOSKPFIOLELRSLGYH--GSSYPDEG 2052
Db 3794 TSDHEVSESRVYASGSPVPSSENSNRVTETTTTTVTRHEHEDEDDQEHVSESQESASG 3853
Qy 2053 VERVPVSSPSLHDKCLPKHLE-----ELDKHLEGLERPKQGPVKGGBAHLHLR 2107
Db 3854 ---SPVPS-----EKSVETVETTTTTVTRHEHHE--DEIPTIV-----ESH-- 3893
Qy 2108 PLPEQSSPPLQTAGVKGHORVWTLAOSHIVITQDYTRHN---POOLASPLPAPLY 2164
Db 3894 ---DDQAASS--VPSEEDVAGIQITTT-----TTTTVTRHVAVDEEIDSGRMELE 3940
Qy 2165 SPFGASCPVLDLRPPS-----DLVLPDPHGAAPRAGSPHSEGGKRS 2207
Db 3941 KYSGES--PV-----PREDESRVIEETTTTTTTRHEHPEDEDDSHVVSQESYASG--SP 3992
Qy 2208 EPNKTSV 2214
Db 3993 VPSEKSV 3999

RESULT 23
T02345
hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Sanders, E.; Robinson, D.;
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
Submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: 214664
A:Accession: T02345
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Reads: 1-1791 <RIC>
A:Cross-References: EMBL:AC004493; NID:g2996648; PTDN:AC08453.1; PTD:g2996650
C:Genetics:
A:Map position: 16
A:introns: 1610/2; 1706/2
A:Note: KIAA0324

Query Match 3.28; Score 416.5; DB 2; Length 1791;
Best Local Similarity 20.68; Pred. NO. 1.9e-07;
Matches 406; Conservative 258; Mismatches 764; Indels 545; Gaps 100;

```

```

Db      1828 FRSSMLPTQMHNL-----ALNRDLTSSVSEKSGCKTP-----GDFNYAYQKP 1871
Qy      1933 MEVLLPREA-----PRVAPRRPADTGHAH-----LAKPPAR 1966
Db      1872 ENAAGSPDEEDYDESOKERTIRTHDVVYVYKTEKRTIKSPDCDSGYSEIETKTKPED 1931
Qy      1967 SGL-----EPASPSKG-----SEPRVLPVSGHATARTAPKXNLAAPH-----2005
Db      1932 GGYTCETKTKTRTPBEGGYEISEKTRTPVSGVYKTEKRSRLLDISNGYDDE 1991
Qy      2006 ---HASPDPAPAPASADPHREKTSKPEFSIOELELSLGYH-----GSSYSPEG 2052
Db      1992 DCGHTLGD-----CSYVETTEKITSFP-----ESESYSVETSKTRSPDTSAYCET 2040
Qy      2053 VEPVSPVSSPS-----LTHDKGLPKH-----LEELDKSLHEGLARK--QP 2091
Db      2041 MEKITKTPQASTVSYETSDRCYTTTEKSPSEARQDVLDLVSSCEFKHAKTELSPSFLNP 2100
Qy      2092 GPVK-LGGEAAHPLHLRLPESQPSPLQI--APGVKHQVVTTLAQHISEVITQDYT 2148
Db      2101 NPLEMPAGE-----EFTESB--KPLTSGGAPPSGGKQ--QGRQCDETPPYSVS 2147
Qy      2149 RHHPQOLAPLPAPLYSPGASCPLY-----DLRPPSDLYL-----PPDHGA 2192
Db      2148 ESASQDTSQDVPEP-----ECPSTADANIDSESEITPTDKTVYKHMDDPP--A 2199
Qy      2193 PARSPHSEGGKRSPEPKTSVLGGEDGI-----BPVSPREGMTEPHGSRSAV 2241
Db      2200 PMQ-----DRSPSPRHPRDVSMDPDALAVQNLGKAVKDKLKEKTKPKTKSS 2251
Qy      2242 YPLLYRDEQTEPMSKSGKSPGNTSQPAPFSKLTESNAWKSCKOINKLNHNNE 2301
Db      2252 SPVKKGDKK-----SKPLAASPFGALKESSDKVRVASPKKESVEKA--TKTTT 2301
Qy      2302 PEYNISQGTETFMNPAITGTGLMTYRSQAVQEHASTMGLAIIKALMGKYDQWESP 2361
Db      2302 PEV-----KATRG-----EKED 2313
Qy      2362 PLSANAFPLNLSASLPAAMPITTAADGRSDHTLTSPPGGGKAKVSGRPSRKASPAQGL 2421
Db      2314 KETKNA--ANASAKSAK--TAT-----TGPYTKTAKSSTVP-----PGL 2350

RESULT 22
T15348
hypothetical protein B0350.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15348
R:Gatung, S.
submitted to the EMBL Data Library, February 1996
A:Description: The sequence of C. elegans cosmid B0350.
A:Reference number: Z18332
A:Accession: T15348
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5170 <GAT>
A:Cross-references: EMBL:U50071, NID:g1208871, PID:AAA93447.1, CESP:B0350
C:Genetics:
A:Gene: CESP:B0350.1
A:Introns: 48/1; 503/3; 5116/3

Query Match      3.2%; Score 419; DB 2; Length 5170;
Best Local Similarity 17.5%; Pred. No. 4.8e-07;
Matches 516; Conservative 371; Mismatches 1020; Indels 1040; Gaps 129;

Qy      16 EPRYP-----HSLSPV-----QIARTHTDVGLLGXGHSRDYASHL--SPG 56
Db      1345 DQFSPAPSSHVSGIHASESPVAQQQEIPTQRFPHEDSPAQY-FHEEYESHVLTEQA 1403
Qy      57 SIIGQRRRP-----SILS 70

```

```

Db      1404 PLTTEQHPESPESGDESGEGLSKVLGFAKKAQVAGVAAVAPALAAVAKAAYDALKK 1463
Qy      71 EFPQGNERSQELHLPRESHSLPRLGKSEMEFIESKRPRLLELPDLR-----119
Db      1464 DDEDEQERESLRLQERSIDSPH--ASBOSQIEEHERFEESVPSEKHHVTTTTTT 1521
Qy      120 -----PSPLLA-----TGQPAGE-----DLTKRSLTGKLEVPSPPHPTDP 158
Db      1532 VTRFVDEHEPLVQELIEGEGKSPADSEKSLPHVVETTTTTTVTRFPKNDSESPVSEK 1581
Qy      159 ELELVPRPLSKBELIQNMNDVDRBITWVEQOISLKLKKQOOLEBAKPPPEKVPSP 218
Db      1582 EOE--PTVSRFVYEAABGDEPEHNYETTTTTVTRKVIDDSQEMGDDDP--KQESQV 1637
Qy      219 IESKHSILVQIITYDEN-----RKAAEAARI-----LEGIG 249
Db      1638 VETTTTTTSREYDNDDETREBAGDSHITETTTTVTRFPHGQEPETETDEVELP 1697
Qy      250 PQVE-----LPLYNPSD-----TRQYHNTKINQAMRKULY 283
Db      1698 PKIEEDNVSEYSSSTSVSREVRPDEPHIETTTTTTVTRFHNPEPETYDDCKDAPI 1757
Qy      284 FKRNHAKQWKQRCQRYQMLEKQVRIENPRRAKESKVRTEYKQPEIR-- 341
Db      1758 SFSQEH-QDDDSQASHQHDR--ESPVESEKSVKHTTTTTTTVTRQLYDEASBIRGE 1814
Qy      342 -----KQRELOERNQSVRGGS 359
Db      1815 SPVATEEHEHVSTKSDSESHVPSVETTTTTTVTRFYDDDELQREDHTQSEERS 1874
Qy      360 GLSMAARSEHEVEEII-----DGLSEONLEK-QMRQLAVLP-----P 397
Db      1875 --SIPTETHEHDHLKETTTTTTVTRFDEPENVKHQDSQFSLSPSSHVSEIYVP 1932
Qy      398 MLYDADQRI-----KFINNGLMADPMKYVKQRYNNMSEQKE 438
Db      1933 ESPVAKQOEIPTQRFPHEDSPAQYFHEDEHGVPTQCALPLTQEQHQ--PESGEESDGE 1991
Qy      439 TPEREKFMQHPKNGPLGLASFLEKTVABCVL-----YYLTGKN-----ENKSLVR--RS 486
Db      1992 GFGSKVLGFAKKAQVAGVAAVAPALAAVAKAAYDALKKEDDEDEDESESLIRERS 2051
Qy      487 YRRRGKSGQQQOQOQOQOQO-----QOQOQQPMPSRSGQEKDE-- 524
Db      2052 FDSPHASQSOILEKHKRFESVPVSEKHNDQSALPQESVQILEKSRFTFNDESERGV 2111
Qy      525 -----KEK-----EAKKEKPE--VENDKEDLKEKTDITSGED-- 557
Db      2112 KSDHYTEDDQSLKPKESGEAFSQFTSEKEQDRSDSPIHQKED-ISOQNSSPEPVK 2170
Qy      558 -----NDEKEAVAKGKRTANSQGRKGRITRSMANSESEAITPQOSASLAMELNES 612
Db      2171 SEQPHDEKPLEROGSYSSGYSFKSPGSGITGL--DEKALSGVDEPRDRENPRES 2226
Qy      613 SRWTEEMETAKGLIEHGR-----NMSAIAVMGSKTVSOCKNYFYKKRQNLDE 664
Db      2227 HEKTE--ATSDENLFEBSDKTAPASPVPSSEDSNRVIETTTTTVTRHFPEDDHSYV 2283
Qy      665 ILQGHKLC--MEKRNARRKKKGAAPAA-----SEEAAPPVED--EEMESGVSGN 713
Db      2284 ESQEVSSSGSPVPEKSVDRVIETTTTTVTRHFEADELPTIVSSHDDOAAFPV-S 2342
Qy      714 EEMVEEAEAL-----HAGSNVPRGEGSPATVUNSS 746
Db      2343 EEDVHQIQTTTTTTVTKHFVDPDEIDSEHNMESDKYAGSPVPSSEDSRVYETTTT 2402
Qy      747 DT-----ESIPSPTEAAKDGTGQNGKPPA-----TL 773
Db      2403 TTVTRHFPEDDHSFPVQTOEYASBPVSEKSVKRVITTTTTTVTRHFPEDDHL 2462
Qy      774 GADGPPPG-----PPTPRRTSRADIEPT--PASEATGAPT-----PP 810
Db      2463 GQGGESDQIPSESITSENMDETSSSPVQSNRDEEFVLPAIAPVKKQPTFGRAVDSHADP 2522

```

F:1861-2064/Region: 17-residue repeats
 F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: pH
 F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (thr) (cd
 F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 3.2%; Score 424.5; DB 1; Length 2464;

Best Local Similarity 19.3%; Pred. No. 1.4e-07;

Matches 521; Conservative 314; Mismatches 996; Indels 869; Gaps 114;

```

QY 68 LLSFPGNGERSQGLHLPESHSLYPLG-----KSEMETESKRPPELLLPDL 118
DB 174 LLSFTHPANKASLTL-----FCPEEGDWKNSLNDRHNLQDFINIKLNSASLIP 221
QY 119 RPSFLATGCPAGSEDLTKORSLGKLEVPSPSPHPHDPLELVP-----RLSKR 170
DB 222 -----EMEGLSSEFTETLS-----ESVESPFPDI-----LEPPTSGFLKLSKPC 262
QY 171 -----ELIQMDRVDREITVEQOI----- 190
DB 263 YIFPGRGDSALFVNGFPNMLINGSEKSCFWKLIRHLDRVD-----SILLTHIGDDMLPG 319
QY 191 --SKLKKKQOOLSEBAKPEPEKVPSPPIESKHSLSVQIITYENRKGAFAHRIIEGL 248
DB 320 INSMLOKRIAELEBE-----RSQSTSNSDMMKXLI 350
QY 249 GPQVELLYNQPSDTROYHENIKINQAMRKLLIFYFRNNAKQWQKQPCQFQDQLEMA 308
DB 351 SPDLGVPLFVNLKQDPEPNIKMKRSTIEACFTL-----QYLNKSMKEPPLFRS 401
QY 309 LEKVE-----RIENPPRRAKESKREYVEKQPEIRKQK--ELQERMQSRVQ 356
DB 402 VGNTELEVLIFQKNQVGLKEMVYVLPVSKSKEMQYFMQWGTGKDAEELLPRNGQEVDI 461
QY 357 RGSGL-SMSAARSHEVS-----EITDGLSEQENLEKQKQULAV----- 394
DB 462 PISYLTSVSSLIWHPANPAEKIRVLFPNGSTQYNILEGLEKXHLDFLQPLATQKDL 521
QY 395 -----IPMLYDAAQOIRIKFINNGLMADPMKYVKQKQVNMMSQOEKE--TFRKPMQHP 448
DB 522 TGOVPTPVKQVVKLQKQK--ADRSRLKPAKTPVASKESKEPELPEVTKTSQVETRP 578
QY 449 KNFGLIASFLERKTAVACVLVYVLTKNENYKSLVRSYRRRGKSOQOQOQ--QOQOQOQ 506
DB 579 K-----VESKQKQV-----LVKKDKPVKTESKQSVTEKEVSKSEQSPVAAEVAEK 623
QY 507 QOQOQPMRPSQOEKQEK--KEKAQKEEKEPEVNDKEDLLEKTDTSQEDNDKEANA 565
DB 624 QATESKPKVTQDKVYKKEIKTKLEKKEKPKK--KEVYKEDKTPLKQDEKPKKEVK 679
QY 566 SKGKKTANSQGR-----KGRITSMANENSEEAITPQOQAEIASMEIN 610
DB 680 KEIKKEIKKEKREKLKEVKKETPLDKAKKEVKKEKKEVKKEK--BEKKEIKKISQDIK 738
QY 611 ESSRWTEEMETAKGLLEHGRNMSAJARWVSKTVSQCKNFYNYKKQNLDELLOOK 670
DB 739 KST-----PQSTTKP-----SALKPKYAKKEESTYK-----EPILAAGK 772
QY 671 LKQEKERNARRKKK---APAASEEAAPPVVEDEMEASGVSGNEEENVEEAALHA 726
DB 773 LKQDGKVKVLEKEGKTEEAATAVGTAAVTAVALAAGIASG-----PVKELAA-- 822
QY 727 SGNVPRGEGCGPATVNNSSDTEIIPSPHTAADTQNGPKPATLGDADPPGCPTRP 786
DB 823 -----EBSLWSSPDDL--TKDFEELKAEIDVAQDI-----KQLELELD----- 860
QY 787 RRTSRADIEPTPASEATGAPTRPPAPSPAPPPVVPVPEKEKEBETAAPVE-----EGE 841
DB 861 -----EELKQETQPEAV-----VIQKETEVSQKSAESPBGITTTTGE 899
QY 842 EQKPPAAEELAVDTGAKEEPVYSKTEBAEGPAKGDAAEATPABGALAEKKEG-- 898
DB 900 GECEQETPEEL-----EPVEKQGVDDIEKFEDEGAGFEESSEFTGVEERKAELEBAEP 951

```

```

QY 899 ---GSGRATTAKSGAGFQDSDSATCSADVEDEAGDKNRLLSPRSLTLPTGDPANA 955
DB 952 EEOEDENASGSAHSGHSTEDDESKANADV-----HAKKRESVV--SQDDAAEE 999
QY 956 SPQKPLDKQKQRAAIIPIQVTKHEPPREDAAPTKPPAPPPQNTQPESD----- 1010
DB 1000 DMDVDLEKGAEOG-----EEGGEEDDAEDAREEGEPDDKTEADYMAVADKAFA 1052
QY 1011 -----APQKQ--SSPKCKSRPAP-----ADKEAFAEAQKLPQGPBC 1048
DB 1053 GVTEEQGYLGTSKQGIQSP--SREPASSIHDETLPGGSEEAATASDEENEDQPE 1109
QY 1049 WTSGLPPEVPREVITKASPHAPDSASVAPPGHPLPLGLHDTARVPLP--RPPTISN 1104
DB 1110 FTATSGYTGQSTIEI-----SSPFTMDMNSTPRDVMSETNNETESP 1152
QY 1105 PPLISSAKHPSVLERQ-----IGAISQKNSVOLHVPYSEHAKAVPVPVTGLPLP 1155
DB 1153 SQEFVNITKYESSLYSGEYKPAVASFNGLSEGSTDATQGDKNYASAS--TISPPSS 1208
QY 1156 MDPKCLA-----PSSGVKQEDLSRGAQGPRESIGVTPAQASV 1194
DB 1209 MEEDKFSKALRDAYCSEKELKASAEIDIKVDSERLSP--AKSPSLSSPP-----SP 1261
QY 1195 LRGTALGVPGSGITKGIPTSRVPSDAITYRGSI THGTADVLYKGTIRIIGEDSPSR 1254
DB 1262 IETKPLGE-----KSVNSLTPNEIKVASGEARVSP-----GVYQAVVEHCASP 1308
QY 1255 LDRGREDSLPKGHVLYEGKKGHVLVEGKMSVTQCSKEDRSSGP--PHETAAKPT 1310
DB 1309 ECKTLEVAVSPQSV--TGSAGHTPYQ-----SPTDEKSHLPRVESENAQAVP-- 1355
QY 1311 YDMMEGKVGALISASIEGLMGAIP-----PERHSHHLKEQNHITGISTQGIIPRYVE 1365
DB 1356 -----VSFEFSKADENBRASLSPMDEVPQSESEVE--KVLSPLSPLLGSSESP-- 1405
QY 1366 AQEDYLREAKLKR-----EGTPPPPPPRDITAEAKYTOALGPLKLPAAHEG 1413
DB 1406 --EDFLADSKVLGRBESPEEGKNGKQGFPPDRSPVSDLT--SGVLQODKQEKSTG 1459
QY 1414 LVATVKEAGRSIHEIPREELRHTPELPLADRLPKESGITQGTPLKYDTGASTGSKKDV 1473
DB 1460 FIPKEQFPEKKTSDVETWSSQALALDERKL--GQDV--SPTQID--VSQFSGFKEDT 1513
QY 1474 RSLIGSPGRTPPPPHPLDWMADARALERACYEBSLKSPGTASSGSGSIARGAVIYEL 1533
DB 1514 KMSISEGTVDKSAATPV-----EGVAEDTYSMEGVAASVSTASVATSS--PPEP 1561
QY 1534 GKPRQSPUTYEDHAPAGHLPRGSPVTMRBEPTRLOEGSLSSKASQDRKLTSTPREI-- 1592
DB 1562 TTDDVSPSLAEGSPHSTEVDSLSVSVQTPPTPQETEMSPKESCPRMSISPPDFS 1621
QY 1593 ---AKSPHASTVPEHPPHPISEYHLLRGVSGVDLYRSHIPLADPTSIPRGIPLDAAAY 1649
DB 1622 PKTAKS--RTPVQDHRSE-----QSMGSIIEFGQESPEHSPAMDFS-- 1659
QY 1650 YLPRHLAPNPTLYPYLYIRGYPTDALENQTIINDYITISQGMHNHNTATMAQAADM 1709
DB 1660 -----RQSPDHPLT-----GASVLAHITENGTEV--DY----- 1665
QY 1710 LRGLSP--RESSALNLYAAGPRGIIIDLSQVPHLVPVPPGTPATAMRLAYLPAPQ 1766
DB 1686 ---SPCDIQDSSLSHK-----IPPEEBSYIQDNDLSLISVSQ 1721
QY 1767 ---PSSRHS--SSPLSPGPTHLYKPTTS-----SSERDRDRERDRERREK 1812
DB 1722 VEAESPSTSSAHTPQIASPLQEDTLSDVPPREMSLYASIASSEKQGLGEG--KLSPK 1777
QY 1813 SILTSTTVHEAPTRWRETQSSSGSSSGSGSSSRPASHAHQHSPISTPTQALQ 1872
DB 1778 SDISPLTPRESSPLYSYSGFSDSTSAKETAAAHQAASSP-----PLDAATAAPYG 1827
QY 1873 QRPVSLNHTQMKGIITAVEBSKPTVLASTSTSPVPPAATFPATHCPGLGTLGVPTL 1932

```



```

Oy 807 TTPAPPSAPPPVVPKEEKEETAAAPVEEGBEOKPPAAEELIA----- 852
Db 88 SPPAPSPBP-PDGPAPBSGAGEAE-AGTPPBGQAGTPPBGQAGAPKPDGTAPQBSGT 145
Oy 853 ---VDTGAEKBPVSECTEEAEGBPKAGKDAEAAEATAE-----GALKAKEKGGSGRAT 904
Db 146 KSGADGKPAQDVPKATTAATEARPAASAAPTVPKATTAATVAAGSAPKATDAANAV 205
Oy 905 TAKSSGAQODSDSATCSADEVDEAGGDKRRLSPRSLTTPDPPANA-----S 956
Db 206 TTAQSQSAKATVEYVPPAAAVAAKEAK-----AVTAAAAAPKATAEAKPAPYTS 253
Oy 957 POKPLDLKOLKORAAATP-----PIQVTKVHEPRREDAAFTKPPAPAP 1000
Db 254 PTPSSSAKAPKPLTAASPTASKATAEAKVPATASIMATKV-----TAEAKP-PSPS 305
Oy 1001 PPONLQPSDAPQOQSSPRGKSRSPAPPADKEAPFAEAOKLPGDPQCMTSGLPPEVPPR 1060
Db 306 VP---KATTDTKAVTATAPK-----AGPDVYKPAVAVCAEAKPAPFP-----PRQ 346
Oy 1061 EVTKASPPAPDPSASTAP---PGHPLPLGLHDTARPLVPPRPPTISNPPPLISAKHSV 1117
Db 347 QLPKAAA-AAAPVTGELKPAATAPRPGSP---RANSHVTVTTPPVN---PRAAAATVETA 398
Oy 1118 LERQIGAISQMSVQLHVPSSEHAKAPVGPVTMGLPLPMDPKLAPFGVQOEOLSPRQ 1177
Db 399 ---GAVPKAST-----GTPPAAAPQOQV-----PKAAPVTPPBPQAVPR-- 435
Oy 1178 AGPPESLGVPTAQAESVLRGTALSGVPGSITKGIPISTR-VPSDAITYRGSITHTPAD 1236
Db 436 ---AATAAAAPV-----TPQOPVTKATTTNATPPPPPIPKAATTTTATV- 478
Oy 1237 VLYKGTITRIIGEDSPSLDKGRBDSLKXGHVYIGKKGHVLSYBGMSVITQCSKEDRS 1296
Db 479 ---TP-----QOPIPKA-----GTD 490
Oy 1297 SSGPHEATAPKRTYDMMEGRVGRAISSAIEGLMGRAIPPRHSPHLKQOHIRGSIT 1356
Db 491 AAPP--AVPKASPD-----GRATP----- 509
Oy 1357 QGIPRSYVEAOEDYLRRBAKLKRGCTPPPPPSHDLTEAYKTQALGPKLKPAAHEGLVA 1416
Db 510 -GVPAATDPQK-----PPPTQS----- 527
Oy 1417 TVKAGRSIHETPRELANTPELPLAPRLKSGSITQGT--LKTYDGASTGSKKRDV 1473
Db 528 ---VPSAVTEPKQOPRAAPPSPNEATPAVSPSPBNLK----- 561
Oy 1474 RSLIGSPGRTPPVHPLDVMADARALERACYEESLSKSPGTASSGGSIAAGAPVIVPEL 1533
Db 562 ---SPLEPTP--KVPFLMA-----LTPQPVTAQ-----MTQQL 589
Oy 1534 GKPRQSPLYEDHGAFFAGHLPRGSPVTMRBEPRLQBSLSSSKASQDRKLITPREIA 1593
Db 590 AATKPSPI-----VPKASPKALMTPPPP-PPGLPRALAAALGLGSPSP--VA 634
Oy 1594 KSPSHSTVEHHPHRPSIEHILRGVSGVDLYRSHPLAFDPDISIPRGILPDLAAAAYIYPR 1653
Db 635 SAMAKATVP-RPLPASP-----VPMASPASISLP--DAARV----- 667
Oy 1654 HLAAPPTYPHLYPPLIRGVDTALENRQTIINDYITISQOMHNHTATAMQRAIDMLGL 1713
Db 668 ALATTAASPGAKP-----EAGGNGTLMAPMG----- 694
Oy 1714 SPRESSLALNAAAGRGITLDSQVPHLPLV---VPPRTGPTATMDRLAYLPTAPQPPSS 1770
Db 695 ---AANTQMAPIGAAGAAQTAPMGAANTHVSPMAGAGTQW-----SPTGAAN 739
Oy 1771 RHSSSPSPGCGPTHILTKPTTSSSERERDRDRDREREKSIIITSTTYEHAPIMRPG 1830
Db 740 TH-MSPIGAGGATQWS-PMGAANTQMSD-----MGATTTQMSPMGAAA 780

```

```

Oy 1831 TEQSGSGSSGSGGSGSSRPASHSHAHQHSPIRPTQDALQORPSVLAHTMGKIITAV 1890
Db 781 TTQSGPM-----GAAATQVATSAAGNTMQVSPMAATP--PQTSV-----GAATTP 825
Oy 1891 EPSKPTVLRSTSTSSPVPRPATTPPPATHCPIGLGLDGVYPLTLMELVLPKAPVAREER 1950
Db 826 QPS--PMGAATTLMSPMGAATTPQPS--PMGAV-----TTQPP--PMATVITQPP 871
Oy 1951 PRADT-----GHAFLKPPARSGLEPRASSKSEPRPLVPVSGHATITARTAKNLAP 2004
Db 872 MAASIPQSTPMGAATTTQSPPMGAATTTQSPPMGASTPQ--AP-----TVAGSP----- 918
Oy 2005 HHASPPDPAPASADPHREKTQSKPESIOELRLSLGHSYSPEGEVPSVSPSPSL 2064
Db 919 ---TPPPIPPSPTA-----QTSQPMSS-----KSPPPDP-- 945
Oy 2065 THDKGLPRHLELDKSHLEGELRPQOPVKLGBEAAHLPHLRPLPESQPSSPILQ7AP 2124
Db 946 ---PKASGAAQTSPPAAHVAN-----ASP 966
Oy 2125 GVXGHQRVVTLAQHISEVITQDYTRHHPOQLSAPLPLPVLVFPQASCVPULDRRPPSDLY 2184
Db 967 GV-----TAVSPAT----- 976
Oy 2185 LPPPDHGAAPRSGPSHSEGKRSPEPNKTSVLGGEDGIEPVSPPEG--MTEPGHSRAV 2241
Db 977 -----GVTEASPSADARLSPTPTAAT-----DG--PASPAAATADVTEAATDV7AA 1021
Oy 2242 YPLLYRDEQTEPFRMGSKSPGNTSQPPAFSKLTESNSANVSKKQEIKNKLTHNENE 2301
Db 1022 ATAAPAEAPKAPKARSSSSSSSSSS-----SSSSSSSSSSSSSDSDSSSSSESNPAS 1075
Oy 2302 PEVNIHQGTETFMMPAITGGLMTYRSQAVQEHASTMGLEAIIRKALMGKYQWESSP 2361
Db 1076 P-----APAVG-----DGOQMT 1088
Oy 2362 PLSANAFNPLNLSASLPAAMPITADGRSDHTLTPSGGGRKAKYGRPSRKAASPAFL 2421
Db 1089 PGAAQSVPTVEAIVQEAHAAAAAAG-----AHEGRTRTRKRRKTRSSSS 1134
Oy 2422 ASGDRPPSVSSVHS 2435
Db 1135 SSSSSSSSSSSSSSS 1148

RESULT 21
QMSPI
microtubule-associated protein MAP1B - mouse
N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei
C:Species: Mus musculus (house mouse)
C>Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 01-Sep-2000
C:Accession: S07549; S44387; A33645
R:Noble, M.; Lewis, S.A.; Cowan, N.J.
J:Cell Biol. 109, 3367-3376, 1989
A>Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A:Reference number: A33645; MUID:90094539; PMID:2480963
A:Accession: S07549
A:Molecule type: mRNA
A:Residues: 1-2464 <NOB>
A:Cross-references: EMBL:X51396; NID:952999; PIDN:CA435761.1; PID:953000
R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Arch. Biochem. Biophys. 310, 428-432, 1994
A>Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A:Reference number: S44387; MUID:94234720; PMID:8179328
A:Accession: S44387
A:Molecule type: protein
A:Residues: 653-663, 'IC' <SAN>
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein; tandem repeat
F:589-786/Domain: microtubule binding #status experimental <MTB>
F:589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69
R:K-E/D-X)

```

```

Db 434 GAVSNQYVSHTPGAALVNLALRPSPHHTFLGHSHPGNIATKXAISEHLKIQFNS 493
Qy 1107 -----PLISSA-----KRPVLERQ-----IGAIQGMS-VGLAHV 1135
Db 494 QAGVFEALPLHKNGFCEDEVHYSHDKVLSGDLQYKVEFYFILINGVAGKGAEEYE 553
Qy 1136 P-----YSEHAK-----APVGPVLMGLPLMPDKPKLAPSGVQEOQLSP 1174
Db 554 PEETPARAYSEGKEMSFDESHIDVDRKSTSPEDNEVPMVSDR-----DSQSLTS- 607
Qy 1175 RGQAGPESLGVPTAQE--ASVLRGTA-----LGSVPGG 1206
Db 608 -----EPPEDELLPDPADPHVMEYVEKEAEEDDEARSQOSQVKEPPEPPEPMSSVPLM 664
Qy 1207 SIKGISTVPSVSSALTYRGSTHGTPTADVLVYGT-----ITRIIG----- 1248
Db 665 DPPKRGPRGPRPKNGIMSKEERLRKKQAMELAKNOPPOPGRGPRKRGPRKPLP 724
Qy 1249 EDSPPRLDRGRDESLPKGHVYIEGKKGHVLSYEGGMSVTQCSKEDGRSS--SGPPEHTAAP 1307
Db 725 EDAPDRPEKKGKXKRRK-----NGEBGDASDAE-----KTIKREKREKTPLELRRE 773
Qy 1308 KRTYDME-----GRVGRALISSASIEGLMGRAIPPERHSPHMLKEQHIRGSITQGIPIR 1361
Db 774 DYTEBQLOKPNKXVGLIDEVLSAAPDGLTKQI-----YKRIQKPYFYFNVDTKGMES 829
Qy 1362 S-----YEAQDYLRREAKL-----KREGTTPPPPPS--RDLEAYTQOLGP 1404
Db 830 SVRNHLIGNDAPKKNBETHLMRSVPGIDIDAGKKRKAPEPDHASLNFQGHVAPQPM-- 887
Qy 1405 LKLPKPAHEGL-----VATKAGRSIHIEIPREELRH-----TPE 1438
Db 888 -----PHRGMVHGHGVQOSHYHPTGVQOROSVYTTQGPAGASQH--PQHLQTPPGVPIPO 941
Qy 1439 LPLAPRLKESGITQGTPLKYDTGASTGSKKHVRSLGSGRTF-----PPV----- 1487
Db 942 QP--FRBAYQAAQTSPPAQ-----AQYGYTPPTAARQMSGTPATYSSPVSRPMTVA 995
Qy 1488 -----HPLDVM--ADARALERACYESLSKSPOTASSSGGSITARGA----- 1526
Db 996 AOSGATPMSMARQSLSPVSGPSQANGIPRV-----NPPATANTGPVAVGAGRAPAOA 1048
Qy 1527 -----PVIYVEL-----GKPROSPLTYEDHGAPFAGHLP-----RGSPTVM 1562
Db 1049 TPTANTTAAPRALNVIAPELISWLESFKVYVEKLEYIQTISKFPQIILAMSVINGLKLT 1108
Qy 1563 REPTP-----RLQ-----EGSLSSRXAS- 1580
Db 1109 KSMIPDESELVVLVRFVEERIQGTSHKSLDPLDQLTLTFKATWSTLEAKLDSQKAE 1168
Qy 1581 -----QDRKLT-----STPRE-----IAKSP-----H 1597
Db 1169 LVLSAIDQVLGLADKTTTGTSEMEREFNNAEKYLIPALRMKVAEMORKOVAATPAAPVH 1228
Qy 1598 ST-----VPEHNPRIISP--YEHLLRGVSGVDLYRSHIPLAFDPTSIPIRGILDAAYY 1650
Db 1229 ATATRGALPLANH--HTMAPATAPSAQAQRIATPSAIGRANPCGANSAPRTITTAAPRP-- 1285
Qy 1651 LPRHLADNPYPLHYRGPYLLRGYDPTALENRQTI-----INDYITSQOMHNTATAMA 1704
Db 1286 -----TPAPILPRAP--ISGHDIAPSTVNTNNAAGARVPVPAVGPQIGVHSGAIG 1337
Qy 1705 QRADMLRG-----LSPRESSALNTAAGRGITDLSQVH--LPV----- 1742
Db 1338 VTPAVSSPPTVTHGVTAAPTPTTTER--SNLAASVAG-----SMAQQTHSGAPAGSGNASI 1391
Qy 1743 -LVPTTGTATAMDLAIVLPTAPQPPSSRHSPLSPGCGTHLTKTPTTSSSERERDRD 1801
Db 1392 SRAPSPATAPTA-----PTSVPVSSYTRPMHSYVPTGPRGILTAFTASSG----- 1446
Qy 1802 REPRDREREKSIITSTTVEHA-----PIWRPTEQSSSGSSSGGGGSSSRPA 1852
Db 1437 ASAGYAPANASTWPAPTSGQAAMTSAVQSVSPRP--SILTTQOMGVPAITAAASTSRPA 1495

```

```

Qy 1853 SHSHAQHSPISPRTODALQORPSVLNMGKGIITAVE-----PSKPTVLR--STSTSS 1905
Db 1496 SGVYNPASSIABSTHSMK--PSAVPTTASGVSSVSSSLATPLPSPAPRGPSNNAT 1552
Qy 1906 PVPRATFPPTATCPLOGTLDGVYPTLMEPVLLPKENAPVARBERPRADTGHAFAPRA 1965
Db 1553 PTTPTATIPRA-----PL-----PAASVSAPVYGPQ-----SFSA--PA 1585
Qy 1966 RSLGPASSSPSKSEPR--RPLVPVSGHATITATPAKNLAPHNASDPAPRAPASDPHR 2023
Db 1586 SVPTPTPTSAASQAOPLSQCPGAPPAISSIGATPPAASI-----PSSA----- 1628
Qy 2024 EKTQSKPFSIQEELMSLGVHSGSYSPGVEPRVSPSSPSTLTDKGLPKHLELDKSHLE 2083
Db 1629 -----PALAPVT----- 1636
Qy 2084 GEIRPKOPGPKVLGEMAHILPHLRPLPE-----SOPSSSPLQTPAGVGHORVVTL 2135
Db 1637 -----YVPQOASAAARLP--VTPAPAAHTIAQSVAPVAPRVTPQS-----VQSV 1681
Qy 2136 AQHI--SEVITQDTRHHPQOLSAPLPALYSFPGASCPLYLDLRRPPSDLYLPPDHGAP 2193
Db 1682 AQHVTSSQAST--TAAP--VAQSVPRPVNSNPTSAAPV-----A 1717
Qy 2194 ARGSPHSEGGKRSPEPVKTSVILGGEDGIEPVSPPEGEMTERGHS--RSAYVPLLYRDEQ 2251
Db 1718 AVGAQVASAPRYTQPPRHALLSVSQSL--POSVPAAQAAQIOTPHSASRPV-----PQ 1771
Qy 2252 TERSRMGSKSPGNTSQPPAFPSKLTBSNSAMVSKKOEINKKUNHNRNEPEYNI SQPQT 2311
Db 1772 SVFQSVQATPQAVPR-----STSLATPT-----AQGP 1801
Qy 2312 EIRMPAITGTMLTTRSQAVQEHASTMGLKALIRKALMGKTDQWESRPLSANAENPL 2371
Db 1802 V--SPAVSGSGVPA--PSAAQSVAPAPV-----SSTPVPAATVAPA 1838
Qy 2372 NASASLPAAMPITRADRSHTLTPSCGGGKAKVSGRPSRKAKSPAPGLASGRPPSVS 2431
Db 1839 STVAAAPTPTRVTAAPALAS-AATNP-----APVSPQ--OHOITGQAP--AQOORPPAOA 1889
Qy 2432 SVHSEGDGCRRTPLTKRWEDRPSASGSTRFPYNPILMRLQAGVMASSPPPGIAPAGSPL 2491
Db 1890 QAPA-----TPTTISAAPRRPPTLAPPPPPPP-----PTEDEPPPPPPPPAPAP 1936
Qy 2492 AGPHHAWDEEPKPL 2506
Db 1937 PPP-----PTPLM 1944

RESULT 20
T18535
high molecular mass nuclear antigen - chicken (fragment)
C/Species: Gallus gallus (chicken)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
Accession: T18535
R/Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A/Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chicken
A/Reference number: Z18955; MUID:9803440; PMID:9365273
A/Accession: T18535
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1151 <SH1>
A/Cross-references: EMBL:DB8440; NID:d1177138; PID:d1025045; PIDN:BA424137.1

Query Match 3.3%; Score 436; DB 2; Length 1151;
Best Local Similarity 19.7%; Pred. No. 2, 4e-08;
Matches 346; Conservative 154; Mismatches 548; Indels 706; Gaps 77;

Qy 752 PSHPTTAUKDTGONGPKRP--ATLGADGPPG--PTPPRR--TSRAPIETPSAETGAP 806
Db 31 PRKWPVIAELHAPAAPQPPKWPVIGAPPPPTETPTPEKPTDGADAPKASAEILT--- 87

```

Db	2473	LVG--LIPSK-----	LSGPLG--PESP--	PSDRLSR-----	PPPAT	2505
Qy	1388	PVSRDL-----	TEAYTQA--	IGPILKJKAHEGL-----	VATVEAGRSIH--	EI 1428
Db	2506	PSSMVNSQVLGSGOAFYQARAPYRGSLPLQOOOQOOLWQOQOATATISRFAMSAFEST				2566
Qy	1429	PREEL-RHTPELPLA-----	PRPKSGSI	TGOTPKYDTGASTTGSKKHVRS	LIGSPG	1481
Db	2566	POPELGRQALGSPPLGISTRELPGP-----	GEEVPGPAGAPCELELHNVOKJLGGGG			2617
Qy	1489	RFPFVHRLDWMADARALERACYESLSKSRPCTASSSGSIAKGRAPVIVPELGRQSPPL				1544
Db	2618	TPFP-----				2634
Qy	1542	TYEDIGAPFAGHLPGRGSPVWMEPTPRLOEGSLSSSKASODKLTSTPREIAKSHSVTP				1607
Db	2635	SDDPH-----	KLAPBGLR-----	GLAVGGLPPQKASAPAP--	ELANSHLPT--	2674
Qy	1602	EHHPPHISP-----	YEHLLRGVGVLDYKSHIPLAFDPTSIPRGIP-----	LDAAAAYLL		1651
Db	2675	---PHTKGPTLPTGELVWVRPSSSTELGRPN--	PLALEAGKLPCEDEPELDDPDHAKLED			2733
Qy	1652	PHHLAPNPYPL--	YPPYILRGYPTALENRQT-----			1684
Db	2731	DEELA-----	HLGLGVDAKDDDEIGTLEMETNDPHLDDLNGDEBDLAAYTDPELDT			2784
Qy	1685	---IINDYTSOQMHNTATAMQORDMLGSLPRESSLALNVAAPRGITLDSQVPH				1733
Db	2785	GDKOITFBNHLVVE-----	SANEERERELNLGVP-----	QPLG--		2822
Qy	1740	LPVLVPPTGPATAMDRLAYLPTAPOP--	SSRHSBPLSPGPTHLYTKPTTSSERE			1797
Db	2821	-PEERPP--	PAADASERPLASVLEVKPKVEEGGRHPS-----	PCQGT--		2860
Qy	1798	RDRBERDRERKESILTSTTYVHAPL-----	WAPG-----	TEOSSGSGSSG		1842
Db	2861	---IATPKVAPAAANSIGLKLPGQSGMGSRDTRMGTGPFSSSG				2902
Qy	1843	---GGGSSSRPASHAHQHSPISTRQDALQORPSVLHNTGMKGITAVESKRTV				1897
Db	2903	HIAEASISGATGPPAH-----	---LITSPISPGSGSSLEKFP-----			2939
Qy	1898	LKSTSSPVREPAATFPRAHCPGLGTLGDVYPLIMEVULLPKEAPYARP--	ERRPADT			1955
Db	2940	LESGLALTPGCPAAS-----	GDELD-----	KMESSIVASEPLILIEDLHEKKE--		2984
Qy	1956	GHAFIAKPPAASGLEPASPSPKSGSPRPLVPPVSGHATTIARTPAKNTLPHNASDPDPAP				2015
Db	2985	---LQKKOOLSAOLOPAAOQOQOQOQOHSILP-----	---APGPA			3017
Qy	2016	ASASDPHEKQSKPFSIOELELRSLIGHGSSYSEGVPEVSPVSSPSLTHDKG--	PHHL			2074
Db	3018	QAMSLPHEGSSPSLGAQOOL--	SLGJ--	AVARQGLPQPLMPTOPRPHALOQRLABSWA		3073
Qy	2075	EELDKSH--	LEGBELREKOPRYVULGGEAAHLLPHLRLEPESOP--	SSSPLQLOTAPGVKSHOR		2131
Db	3074	MVSNQGHMLISQ-----	HGGQAG-----	LVPQSSQGVLSQKXPMTGMRPMSCKMPQ		3119
Qy	2132	VVTLAQHISEYI--	TODYTRHHPOOLASPL--	PAPLYSPFGAS-----	CVLLDIR	2178
Db	3120	QIAMQOOLANSFFPDTDKRAEDDITGPIAKAKWALKGIKYMAQSGSIGVAGPMNRQQ				3179
Qy	2179	---PSPSLYLRPPRDHGAPARGSHSBGKRSREPNKTSVLGGBEDGIEPVS				2227
Db	3180	VSLLAQRLSGGBSSDL--	---ONHVAASGGORSADSPQPPRPNPTPAQGVINAEADROXY			3235
Qy	2228	PEGMTEPGHRSAAVYPL--	---YRDEQOTEPBSMGKSPGNTSQPAPFSKLTES			2278
Db	3236	EEWLF--	HTQOOLQMOJKVLEBOIGVHKRSRKALCAKQAPAKKAGRFFPADDEKL--			3289
Qy	2279	NSAMVSKKQOETINKKLTNNHNEPEYNTISQPTETENNPATGTGLMT--	YRSQAVQSBAS			2337

[illegible]

Db 2875 VSLLAQRLSGSPSDL-----QNHVAAGSGGERSAGDPSQPRNPPTAQCIVINADDRQY 2930
QY 2228 PEKTEPGEHRSRVAVPL-----YRDEQTEPSRMGSKSPGNTSQPPAFSKLTES 2278
Db 2931 EEWLF--HTCQLLOMOLKYLEEQIGVHRKSRKALKAKORTAKKAGREFEADAEKL--- 2984
QY 2279 NSMNVKSKQKQINKLNTNHNREPEVNIISQGTETFMMPAITGTGLMT--YRSQAVQEHAS 2337
Db 2985 --KLVEEQSKIQKQLQVRRQCKEH-----TNLAETRNKQKQKQ--- 3023
QY 2338 TNNGLAEIIRKALMGKYDQWESPPSLSANAFNPLNASASLPAMPITAADGRSDHTLTSP 2397
Db 3024 -----QQQQQQQQQSHSAVLAISP--SQSPRLITKLPQGLLGC---HGQOP 3064
QY 2398 GGGGKAVKSRPSSRKAKSPAPGLASGDRPPSVSSVHSEGDQNRRTLPTRKWEEDRSSA 2457
Db 3065 QG-----PPGQAGG----- 3074
QY 2458 GSTRPFPNPLIMPLQAGVMSPPRPLP-----AGSGPLAGP 2494
Db 3075 -----LRLTPGGMALPGQPGPFLNTALAQQQQQQSHGAGSLAGP 3115

RESULT 18

T03454
ALR protein - human
C/Species: Homo sapiens (man)
C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 27-Oct-2003
C/Accession: T03454
R/Parad: R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A/Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A/Reference number: Z14954; MUID:97388474; PMID:9247308
A/Accession: T03454
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-5262 <PRA>
A/Cross-references: EMBL:AF010403; NID:92358284; PIDN:AACS174.1; PID:92358285
C/Genetic:
A/Map position: 12
C/Superfamily: acute lymphoblastic leukemia protein, ALR type
C/Keywords: alternative splicing

Query Match 3.6%; Score 470.5; DB 2; Length 5262;
Best Local Similarity 20.3%; Pred. No. 7.6e-09;
Matches 558; Conservative 284; Mismatches 891; Indels 1019; Gaps 144;
QY 145 LEVUSPSPPTHTDELELVPRLSKE-----LIONMDVDEITVNEQISKLK 194
Db 1286 VKPVAVAP-----ELVPMKVKEPPOYFRFGVWLTEGMALNLTN----SPLH 1334
QY 195 KKOQO-----LEEEAA-KPEPEKPVSPPIESGHSILVOIYDENRKAARHILIG- 247
Db 1335 KRRQRRLGLPGAGJEGSEPSDALGPDKKODLDTDELKSGGVENHECIIKJEGP 1394
QY 248 LGPQVELPLYNQSDTRYHENIKINQAMRKL-----ILYKRRNNAKQWKQKFC 299
Db 1395 VSPDVE--PGKEETESKK-----RKRKPYRPGIGGFWMQRKSHTRTK-KPPAA 1441
QY 300 QO-----YQLMALAEKXRYETIENNRRAKSKVAYEYEQ 336
Db 1442 QAEVLSDGQPDDEVITPADLPBEGAVESQSLAEGDEKKQO-----RRRKRKSKLEGMPAY 1496
QY 337 FPEIRKQRELOERMQR-----VGQRSGLSMSAAREHEVSEIIDLSEQENLEKQMRQ 391
Db 1497 LGAEFFKEKLLDLRSKRLFAVGVGRPSFGIGTPKAKG-----DGSSEKELPTS--Q 1546
QY 392 LAVIPMLYDADQORIK-----FINNGMLADPMKYKQORQVNNMMSSEQEKETFRKFM 445
Db 1547 KGDGDPDIADBEBSGLEGKADTGPEDGQVKAESP-----SDPEK----- 1587

QY 446 QHPKNFGLASFLERKTVACVLYYLLTKKNENYSLVRSYRRRKGQQQQQQQQQQQ 505
Db 1588 PGTPEGMLSSDLDRISTEE----- 1607
QY 506 QQQQQPMPRSSQOEKDEKEKEKE--AEKEE--KPEVNDKEDLLK----- 547
Db 1608 -----LPK--MESKQQLQPKDVLASEREQHGCGTPGLEGRTPLOQPLQGLN 1659
QY 548 -----EKDDTDSG-----ENDKEAVASGKRTANSQGRKGRITPSMANEANSBAIT 597
Db 1660 LPSSPMDSYVGCQSPFLDSRGGFFSPDEBPDSQPMWGSQTPSTPT-----T 1711
QY 598 PQQSABLASMELNESS--RWTEEE--METAK-----KGLBGRNMSAIAMVG 642
Db 1712 PTTBGEEDGLSYQRSIQRMKDEBELQGLSTIPVLYANINFPRLKODPYDWS----- 1764
QY 643 SKTVSQCKNFYFNKKRONLDEILOHKLKMEKERNARRKKKAPAAASEEAAFPVVED 702
Db 1765 -----SRCKQIMKLMRKVPADKAPVYLQAK--DNRAAHRINKVQKQAESQ-----IN 1810
QY 703 EEMBASGVSGNEEMMEALALHAGNEVPRGCSGATYNNSDTESISPTTEAKOT 762
Db 1811 KQTKVGDIAKTRD-----PALHLRIPOGALGSPRA--AAPTFISGPTTPAGLST 1862
QY 763 GONG-PKRPATLGADGPPGPPT-----PPRTSRA-----PIEP----- 796
Db 1863 SADQFLKPPA-----GSVPDPDSBELFLKLPQVPAQASQDPFGIAPVLEPRPTA 1917
QY 797 -----TPASEATGAPTEPPAPSPSPAP-----PVVAPKEKEETAAAPVVEGE 841
Db 1918 PPTVPPVPSPTGAQAPQPMLGASRPQAGGFFHTTPTGTPRHQSTPDPFLKP----- 1972
QY 842 EQPRAAEELAVDT-----GKAEPYKSECTEBAEGRPKGDAEAEATGALKAEK 895
Db 1973 --RCPSLDNLAVESPVGCGKASBPGLSP-----PPGSRK-----ALEVK 2014
QY 896 KEGSGRATTAKSSGAPQ-----DSDSAT-----GSADVDEAGDKNRLSPRSL 945
Db 2015 BEIGA-----SSPYGPRNLGFVDSRSSGTHLGLIELKTPVFKAP-----LTPRSQV 2063
QY 946 TPTGDPANASPOKPLDKQKORAAAIPIQVTKVNEPRRED-----AAETKP 994
Db 2064 EPQ--SPGLGRPQEP-----PPAQALAPSPFHNDIIRPGSYTDPAQCPRLT 2109
QY 995 APPAPPPPOQ--LQPSDAPQOPGS-----SPGKSRSPR-----PADKAPAAEAKLP 1043
Db 2110 PRQPPPPPPSCALPPRS-LPSPDFSRVPVSPQSSQSPGLTPRPLSAAFC----- 2161
QY 1044 GPPPCWTSGLPFPVPPREVIKASBPADPSAFSAPRGHPL--PLG-LHDTARVPLPR 1099
Db 2162 -----PSVTPPRF-----QSPDPYS--RPPSRQSRDPFAPLH--KPRPQ 2199
QY 1100 PTISNPPLISSAKHPSVLEROIAGISQGSVOLHV--PYSE-HAKAPVG-----PVT 1149
Db 2200 PEV-----AFKAGSLAHTSLGA--GGRPALLPAGPAELAKVPSGQPPNVRSRPGT 2249
QY 1150 --MGDLPM-----DPKTLAPRSGVKQEOQLSP-----RQAGPSELGVPTQABAS 1193
Db 2250 GAFVGTSPSPKRTFPQAVGEBSLKP--VPQGLPFPHGINSHFGGPILKQOSTNYT 2306
QY 1194 VL-----RGALGSVNGGS--ITKGIPSTRVPS-----DSAITYGSITHGPRADVLY 1239
Db 2307 VATGNHPPSGSPGLPSSGSGTESYGLSPPLPSPVLPAPADGSLPY--LSHGASQ--R 2360
QY 1240 KGTITRIIGDSPRLDRGREDLPKGVHYBEGKKGHVLYEGGMSVTQCKED----- 1293
Db 2361 SGITSPVEKKEPDG--TGWGSILATAEL--PGQDPGMS--GLSGTELEKQORQRLR 2412
QY 1294 -----GRSSGPRPHETAPKRTYDMEGRGRIASASIF-----G 1329
Db 2413 ELLIRQOIQNTLROEKETAAAAAGAVGPPGSGAEBSSPAFQOLSQGTQPPAGTDOKSS 2472
QY 1330 LMGRAIPPERHSPHHLKEQHIRGSIQGIIPRSYVENAEOBDYLREAKLTKREGTTPPP-- 1387

QY 337 PFEIRKORLEOERMOSR-----VCGORSGLSMSAARSEHEVSEIIDJISEOENLEKQMRQ 391
 Db 1192 LOEAFPGELLDLBRKALFANGVGRPSRGIGTPKAKG-----DGSERKELEPTS--Q 1241
 QY 392 LAVIPEMLYDADOORIK-----FINNGLMADPMKYKORQVNNMMSSEOKETFERREKFM 445
 Db 1242 KGDGDPDIADSEESGLEGKADTPGEDGCVKASVP-----SDPEK----- 1282
 QY 446 QHPKNFGILASFLEKRTVAECVLYYYLTKNENYKSLVRSYRRRGKSQOQOQOQOQOQOQOQ 505
 Db 1283 PGTBEGBLSSDLIDISTEE----- 1302
 QY 506 QOQOQOPMRRSSQOEKEKEKEKE--AEKEE--KEVENDEKEDLIK----- 547
 Db 1303 ---LTK--MESDLOQLFQFDVLSGSEBOHGGCTGPLESGSRPLQRPPLQOGLPLGN 1354
 QY 548 ---EKTDDTSG-----ENDDEKAVASKGRKTANSQGRKGRITREMANEANSSEAIT 597
 Db 1355 LPSSSPMDSYPLQCSPPPLDSNERGFPSPERGEDSPWTSQGTTPSTPT-----T 1406
 QY 598 PQOSAEIASMELNESS--RWTEE--METAK-----KGLLEHGRWMSAIAARMVG 642
 Db 1407 PTTBESGDLSTNQSLOMEKDELGOLSTISPLVYANINPNLKODYDPMW----- 1459
 QY 643 SKTVSOCKNFYFNKYKRONLDEILQOHLKMEKENARKKKKADAAASEBAAPPVED 702
 Db 1460 ---SRCKQIMKLMKVPADKAPYLQKAK--DNBAARINKVQKQAESQ-----IN 1505
 QY 703 EEMESAGVSGNEEVEEAEALHASGNEVPRGECGPRATVNNSSDTESTPSPHTEAAKDT 762
 Db 1506 KOTKQGDIAKTDR-----PALHLRI PQQALGSPPPA--AAPTIIGSBTTAGIST 1557
 QY 763 GONG--PKPATIGADOPRPGPPT-----PPRRTSRA-----PIEB----- 796
 Db 1558 SAGDGLTKPRA-----GSVGRPSRGEILFKLPPQVPAQAPSDPPGLAVYLEBRPPTA 1612
 QY 797 ---TPASBATGAPTPPPAPSPSPAP-----PPVVKKEKEEETAAAPVEBGE 841
 Db 1613 PPTVPPYPSPTGAPAPQPMIGASSRPGACQPSFHITPPTGRHOPTSTDPFLK----- 1667
 QY 842 EOKRPAEELAVDT-----GKAEPVVKSECTEBAEBGPAKGDMAAATALEGALKAEM 895
 Db 1668 --RCPBLDLAVAPESFGVGGAKEPBLSP-----PPGSESK-----ALVYVK 1709
 QY 896 KEGSGRATTAKSSGAPQ-----DSDSAT-----CSADEVDEABGDKXNLLSPSL 945
 Db 1710 BELGA-----SSPSYGPPLGFDVPSSTGHLGGLBKTDVFKAP-----LTPRASQV 1758
 QY 946 TPTGDPANASPOKPLDLKOLKORAAIPIQVUTKTHEPRRD-----AAPTKE 994
 Db 1759 EPO--SEGLGIRPOEP-----PPAOLAPSPSHPIIFRPSGYTPRYAOPPLT 1804
 QY 995 AAPAPRPPON--LOPESDAPQOQGS--SPRGKSRSPAP--PADKEAPAAEAKOXL 1043
 Db 1805 PRDQRPSPSCCALPPRS--LPSDPFSRVYVSPQSSOSPLTPRLSLAEATC----- 1856
 QY 1044 GDRPCTWSGLPPEVPREVYIKASPHAPDPSAFSYPAPGRL--PLG-LHDTARPLVLP 1099
 Db 1857 ---PSPVTRRF-----QSPDPYS--RPPSRPOSHDPPAPLH--KPPRPOP 1894
 QY 1100 PTTSNPPPLISSAKHPSVLEKQIGALISQMSVOLAH--PYSE-HAKAPVG-----PVT 1149
 Db 1895 PEV-----AFKASLHATSLGA--GGFPALPAGAGELTAKVPSGQPPNFVASPOT 1944
 QY 1150 ---MGLPLPM-----DPKTLAPPSGVKQOQLSP--RGQAGPPRESLGVPTAOEAS 1193
 Db 1945 GAVVGITSPKRFTRPQAVGERSLKP--VPOGLPRPHGINHFERGRLGKROSTNYT 2001
 QY 1194 VL-----RGTLASVPGGS--ITKGLPSTVPS-----DSAITTSGITTHGTPADVLY 1239
 Db 2002 VANGNFPHPSGPIGSPSSGSTGESYGLPLRPPVLPAPADGSLPY--LSHGASQ--R 2055

QY 1240 KGTITRIIGEDSPRSLDRGREDLSLPGKHIVYEGKKHVLSEYGGSMTQCSKED----- 1293
 Db 2056 SGITSPVEKREDG--TWGSSLATAEL--POTOPGMS--GLSOTLEKQROQRRLR 2107
 QY 1294 -----GRSSGPPHETAPARKTYDMMEGVRGALISSASIE-----G 1329
 Db 2108 ELLIROIQONTLROKETATAAAGAVPPGSGABSPSPAFEOUSRGOTPPAGTODKSS 2167
 QY 1330 LMGRAIPPEHNSHHLEKQHHTIGSTIGCI PRSYVNAQODVLYRREAKLLKEGTPPPP-- 1387
 Db 2168 LVG--LPPSK-----LSGPILG--PGSF--PSDDRUSR-----PPPPAT 2200
 QY 1388 PPSRDL-----TEAYKTOA--LGPLKLPRAHEGL-----VATYKAGRSIH--EI 1428
 Db 2201 PSSMDVNSROLVGGSQAFYRAPEYPSLPLQOQOQOQLWQOQATATSRFMSABAFPT 2260
 QY 1429 PREEL--RHTBELPLA-----PRPLEGSTIGCTPLKYDTGASTTSKXHDVSLGSPG 1481
 Db 2261 PGPELGRQALGSLAGISTRLPGP-----GEPVGPAGAPQFTELRRNVQKGLQPGG 2312
 QY 1482 RTPPVHPRLVMDARALEACVBEELKSRPGTASSGGSIAAGAPVYVELGKPPQSP 1541
 Db 2313 TPFP----- 2329
 QY 1542 TYEDHAPFAGHLPRGSPVTMRREPTRLQEGSLSSKASQDRKLTSTPREIAKSPSTVP 1601
 Db 2330 SEDH-----RLAPBGLR-----GLAVSGLPQKPSAPAP--ELNNSLHPT-- 2369
 QY 1602 EHNHPISP-----YENHLRGVSGVDLYRSHIPLAFPTSTPRGIP-----LMAAAVYL 1651
 Db 2370 ---PHTKGPPLPTGLELVNRPSPSSTELGRPN--PLALEAGKLPCEBDELDDFAHKALED 2425
 QY 1652 PHHLAPPTVPHL--YPPVLLRGYEDTALENROT----- 1684
 Db 2426 DEELA-----HGLGVDAKADDELGTLENLENDHDLNDLNGDFDLATDPDELDT 2479
 QY 1685 ---IINDYTSOQMHNTATANAQORADMLRGLSPRESSIALNYAAGPGIIDLSCVPH 1739
 Db 2480 GDXKDIFNEHRLVE--SANEAEAREBALRGVEP-----GPIG----- 2515
 QY 1740 LPLVLPPTTPATAMDBLAVLPTAPOPF--SSNHSPLSPGPGTHLTPTTSSSERE 1797
 Db 2516 --PEERFP--PAADAEPRLASVLPVKKRVEEGHPS-----PCOPT----- 2555
 QY 1798 RDRDRERDREREKSLITSTTVEHAPI-----WRPG-----TEOSSGSSSG 1842
 Db 1843 ---GGGSSSRPASHASHAHQSPISPRTODALQORPSVLIHNTGMGIIITAVPSKPTV 1897
 QY 2598 HTAEKASFGATGCPRAH-----LTPSPISGSGSSLEKFE----- 2634
 Db 1898 LRSTSTSPVAPRATFPPTATCPIGTLDGVVPTLMVEVLLPKAPRYAR--BRPADT 1955
 QY 2635 LESGALTLPGAPRAS-----GDELD--KMSSLVASELPULIEBLLHNEKKE-- 2679
 Db 1956 GHAFIAPKPPASGLEPASPEKSGSPRLVPVSGHATARTPAKNLAPNHASDPRAAP 2015
 QY 2680 ---LOKQOLASQLOPAAQOQOQOQOQOHL-----APQPA 2712
 Db 2016 ASASDPHREKTOKSPFSIOLELNSLGYHSSYSPEGVEPYSPVSSPSLTHDKGL--PHL 2074
 QY 2713 QMSLPHHGSSPSLAGSQOOL--SLGL-AVAROGRLQPLMPQOPRHNALOQRLAPSWA 2768
 Db 2075 BELDSSH--LEGELARKQGVYKLGGEAHLPHLRLPESOP--SSPPLQTAFGVKQOR 2131
 QY 2769 NVSNQGHMLSQ-----HGQAG--LVPOSSQVLYLQKMGNTMRPMCMKQ 2814
 Db 2132 VVTLAQHISEVI--TODYTRHNPQOLSAPL--PARLYSPGAS-----CVLDIR 2178
 QY 2815 QIAMAQOOLANSFPPTDIDKFALEDIIGPIAKKAVVALKGIKVVAAQSIGAVAGMNRQ 2874
 Db 2179 -----PPSDLYLPPPDHAPAGSPHSEGKSPBPNTKTSVLGCGEDGIEPVP 2227

```
QY 979 TKHERRRDAAPTKRAPRAPPQ-----NLQPSDAPQQP-----G 1016
Db 1008 -----AVPAPPPAPPPERRKCFVLVQASLNRPEALEAVPRENKRAVRPAA 1057
QY 1017 SSPGRKSRP-----APPADKEAFAEAQKLP--GDPKWTG-----GLPFV 1057
Db 1058 SKPTKTSVPOISVCTTQGGSGGK-----SOMQDRPLGSSPPTTALQVTPGLQLP 1113
QY 1058 P-----PREVIKASPHAPDS-----AFSYAPRGHPLPLGLHDTAPVLPPTI 1102
Db 1114 PASLFISQQLLPQOQSSSEFFPTQAMAGLISPSYMP--PLPSLQQA--PLPLQPTV 1169
QY 1103 SNP-----PLISSAKH-----PSULEROIGALSQGMSTVL-----HV 1135
Db 1170 LHPQLHLPLQLPHADIPFOQPPFLMPCRPASTLSGVFLPLQSQPFLQPLPEISHL 1229
QY 1136 PYSEHAKAPVGVPMGLPLPMDPKKLAPFSG-----VKQQLSPRQOAGPESLGVPTA 1189
Db 1230 P---PVKTSLEPLATG---PPGSSSTHEYSSDQLPPTPQATSPAPTSAPP--LALPAC 1281
QY 1190 QEASV-----LRGTALGSV-----PGSITKGIPTSTVPSDSATYRG 1227
Db 1282 PDAMVSLVVPVRIQTHMPSYGSAWYTLISQILVTPQSPASTALTKEEPPSSKMTV-- 1339
QY 1228 SITHTGADVLTKGTIRIIGEDSPRLDRGRBDSLPKGNHYEGKKGHVLSYEGMSVT 1287
Db 1340 -----CEADV-YE-----AEPSPSISKEONRGYOTPYLRVBERKGTSLSSGILSLE 1386
QY 1288 QCS-----KEDRSSSGPPHE-----TAAPK 1308
Db 1387 GCSSTAGSGKRVLPAGSLBLTMTQOQKRVKEBASGADKELVLTCSVVLSTBERK 1446
QY 1309 RTYDMEGRVRA-----ISSASIEGLMGRALPPEHRSPHLKEQHHRIGSTQG--IPR 1361
Db 1447 KTEKPHVGGQGRSRRBAETLSSLDVSDPKELPSLHS-----TLHGTAAPG 1494
QY 1362 SYVNAQDYLAREKALKREBTPPPPPRSLTBAVTKQALGPLKRAHGLVATYKEA 1421
Db 1495 S--BALKEVAPSSK-AHRLPLP-----MSVVK 1520
QY 1422 GRSIHETPREELRTPELPLAPRLKEGSITQGTPLKYDTGASTGSKHDSRLIGSPG 1481
Db 1521 -----EDPKQGTDLPLAPSSSLPLSD---TSPPRAKIQEG---TDSKK-----V 1559
QY 1482 RTEPPVH-----PLDVM-ADAPALERA--CYEESLKSAPRGTAASSGGSTAR 1524
Db 1560 LQFSLHTTTNVSWCYLNVIKPNHIQHADRSSVYAGCISLYPNLPGVSTKALSLIR 1619
QY 1525 GAPIVVELGKPROSPLTYEDHGAPFAGHLPRGSPVYMBETPRLQEGSLSSSKASQDRK 1584
Db 1620 -----SKOKYSKETTYMATAP--HPEAGRLVPSNSRKPMTTEVHLNLSVSPESQK 1667
QY 1585 LTPREIAPKSPHS-----TWPEHHPHPISEPHLLRGVSGVDLVRSHIPLAF----- 1632
Db 1668 ---DPAVKEKEKQKAEGETPTSKGGRPAVKIFE-----GG---YKSMBEVIYVGR 1715
QY 1633 -----DPTSIPIRGIPLDAAAAYLPRHL-----APNPTYPHLYPPYLIRG 1672
Db 1716 GRGRVCECGIRCKKPEMLKXHIRTHDVAPYCKCHFAFKTKGNLTCKMKKASHSK 1775
QY 1673 YPDTALLENROTIINDYITSOQMHNTATAMAOBADMRLGSPRESSLALVYAAGPRGI 1732
Db 1776 COETGVLEELA---EBGTSDDLHOD-----SRQEG-- 1804
QY 1733 DLQVPHLPVLPPTPGPATAMDLAYLTPAQPFSSRHSSPLSPGPTHLTKPTTS 1792
Db 1805 -----AAAVE-----EHQPSDLF-----DS 1819
QY 1793 SGEERDRDRDRDREREKSLTSTTVEHAPLWPRTEQSSGSGSGSGGGGSSSRPA 1852
Db 1820 DSDSDLDDEDEEBEESODELSGPCS--EAPAPCLPPTLOENSGP----- 1864
```

```
QY 1853 SHSHAHQSPISP-RTODALQORPSVLANHGKGIITAVEPSKPTVLSTSTSPVPA 1911
Db 1865 -----VEGQADSTSDVEPEGST-----SEATHLTASSCSTPSQTQ 1903
QY 1912 TFPPTHCPRLGGLTLDGYPTLMEPVLLPKCAPVAREPRADTGAFLAKPPASGLEP 1971
Db 1904 GLPRLGLAPL-----EKDMSAPSPATSPRRMS-----PSKEAGSRP 1942
QY 1972 A-----SSPKSGSEPRVLVPVSGHRTIARTPAKNLAPHNASDPDPPAPASDPNHEKT 2026
Db 1943 SVTRKSHLTKNDSSPOQCSAPARAQASVTSTPGQWQGR-----DLGPHLCSPPBELS 1997
QY 2027 QSKPFSIQELELSLGVHSGSSYPEGVEPVSPSSPSELTDKGLPKLLELDKSHLEGL 2086
Db 1998 RLTPYPI-----GREAPGLERAT-----DTGTPKRSPTRRNSLGAAS 2036
QY 2087 RPKQ--PGPVKLGEEAHLPH-----LRPLPESOPSSPRLIQTAPGVKQHQRVTLAQHT 2139
Db 2037 PQTVLPGKVALAGPCSPSADKSGQLGLGVPRRA-----LLQVPV--LPH-----TLLSRS 2084
QY 2140 SEVTDQYTHHHPQQLSAPLPAPLYSPFGASCPLYDLRRPSDLYLPPDHG--APARGSP 2198
Db 2085 PETCTSAMRKTESRSPSAG--PGPLPRPFSa-----PHDF-----HGHLPSRSEE 2128
QY 2199 -----HSEGGKRSPEPNKTSVLGQ-----GBDGIPEVSP-----EGMTEPGHSR 2238
Db 2129 NLBSHLPLHSLQLSRACP--LIPIGIQVQVAPRGQPTVLPQPCAMVSGFSGGSD 2186
QY 2239 SAVVPLLYRDGEQTEPSRMG--SKSPGNTSQPPAFPSKLTESN 2279
Db 2187 TGA-----REAGERSRMSPTSPSASVSPVAKVSFTLSS 2221
```

```
RESULT 17
T03455
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 27-Oct-2003
C:Accession: T03455
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homology
A:Reference number: 214954; MUID:97388474; PMID:9247308
A:Accession: T03455
A:Status: preliminary; translated from GS/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4957 <PRA>
A:Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: acute lymphoblastic leukemia protein, ALR type
C:Keywords: alternative splicing

Query Match 3.6%; Score 470.5; DB 2; Length 4957;
Best Local Similarity 20.3%; Pred. No. 7.1e-09;
Matches 558; Conservative 284; Mismatches 891; Indels 1019; Gaps 144;

145 LEVPSPSPPTPELPELVPRLSKEE-----LIONMDRVDEITWVEQOISKLK 194
Db 981 VKPVAVPAP-----ELVPMKVKEPEQYFRFEGVWLTEGMALRLNLTN-----SPLH 1029
QY 195 KKQOQ-----LEERAA-KPPEPEKPVSPPIESGHRSLVQIIVENRKKAAMARILIEG- 247
Db 1030 KRRQRGRRLGLPGAGLEGSEPPDALGPDKKQGDLDLTDELKKEGGEVHEHCEIKLEGP 1089
QY 248 LGPQVELPLYNQPSDTEQYHENIKINQAMRKL-----IIFYKSRNHAKQWKQKFC 299
Db 1090 VSPDVE--PGKEETESKK-----KKRKYRPGIGGFVWYRKQKSHTRTK--KCPAA 1136
QY 300 QR-----YDQMLEALEKKVERIENNRRAKESKVRVEYEKQ 336
Db 1137 QAEVLSGDQGPDEVITPADLPAEGAVESGLAEQDEKKQO-----RRGRKRSKLEGMFPAY 1191
```


QY 1791 TSSSERERDRDREREREKSLITTTVEHAPIWRPTEQSSGSGSGGSSSR 1850
Db 2222 PALSE-----GSSSEATTPISSVAREFP---PGLF-----AAEQSAEGISGKE 2263
QY 1851 PASHGHAHSHSPITSPRTODALQORPSVLHNTMGKIIITAVEPSKPTVLSTSTSSPVPA 1910
Db 2264 SAASH-LMDLTPLSAPASLIDLPA-----PA 2290
QY 1911 ATRPATHTPLGTLTDVYPTLME-----PYLLPK-----EAPR-----V 1945
Db 2291 PAPAPAPGLP-GDLQDGLTPCRPECTGELTKRPSPLSPSGHEANGPEITSLNPGFVT 2349
QY 1946 APERPRADTGHAF-----LAKPARSGLEPASPSPKSGEPRLPVVSGHATI 1994
Db 2350 AAKEEBEAPAPAMRGSGWBEAGERSRDTLLSSEQPLRPKSGSG-----PFCALSSEV 2405
QY 1995 ARTP--AKNTAPH--HASP--DPPAPASADPHREKTSKPPSIQLEL-----RS 2040
Db 2406 EAGPOGCATDPRPHCGELSPSFLANPLPSP-----TDDSDLSTEARLAGKGRRR 2456
QY 2041 LGYHSSVSP-----EGVEVSPVSPSLTHDKGLPKHLELDKSHLEGLAPKQGP--P 2093
Db 2457 VGRPATGCPMADETPPTTSASDSSQSDVDPEPTEBCPSTAAALSDDEDGDLF 2516
QY 2094 V-KLGG-----EAAHLPHLEPLPESQSPSLLQTAQVYKQGVVTLAQHISEVITQ 2145
Db 2517 VDKAGVGSGTHHPRGHDPPTPLPDRPSP-----RP 2550
QY 2146 DYTTHHPQOLAPLPAFLYSPFASCVYLDLRPPSDLYLPPDHGADARSGSPHSEG--- 2202
Db 2551 DVCMADEPGLS-----ESGRVERLREKGR-----PGRAPGRAPKAPSPARL 2593
QY 2203 ---GKSPSPKNTSVLGGGEGDIEPVSPPEGMT---EPGH---SRSAVYPLVRDQGOT 2252
Db 2594 DIRKRSPTPGKPYDRISRTVPRTSPSTVSABEKDGHSPMSKGLV-----NGIKA 2647
QY 2253 EPSRMGSKSPANTSQPPAFSKLTESNAGMYKSKQZINKKLNTH---NENEP 2302
Db 2648 GSTALGSK---GSGSPYVVDLAYIPNHGSKGTMDQDFRRVRASYYVVGNDP 2698

RESULT 16

T42717
DNA-binding protein Rc - mouse
N:Alternate names: Ig kappa chain gene enhancer Recognition component
C:Species: Mus musculus (house mouse)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T42717
R:Wu, L.C.; Liu, Y.; Strandmann, J.; Mak, C.H.; Lee, B.; Li, Z.; Yu, C.Y.
Genomic 35, 415-424, 1996
A:Title: The mouse DNA binding protein Rc for the kappa B motif of transcription and for
a family of large transcriptional proteins.
A:Reference number: 222238; MUID:97001141; PMID:8812474
A:Accession: T42717
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2282 <WUL>
A:Cross-references: EMBL:L46815; NID:91377885; PID:91377886; PIDN:AA040884.1
C:Experimental source: strain BALB/c; clone TL; thymocyte, brain
C:Genetics:
A:Gene: Rc
C:Function:
A:Description: binds V(D)J recombination signal sequence and kappa B motif
C:Superfamily: HIV-1R enhancer-binding protein
C:Keywords: DNA recombination; transcription factor

Query Match 3.6%; Score 477; DB 2; Length 2282;
Best local similarity 19.3%; Pred.No. 1.8e-09;
Matches 518; Conservative 288; Mismatches 900; Indels 976; Gaps 127;
QY 7 LVATWATERPRYPHGISYVQIARTHTVDGLLEYOHSHSDVASHUSPSSITIQPORRRP 66

Db 107 LEGSTWQVDMRPDQSGFVA--PGSHPOQL-----PSHAS-----ILPEELP 151
QY 67 SLSEFQGNRRSOEHLRPESHLYLPELGSEMEFIESKPR-----LELLDPLLRP 120
Db 152 GIPKVFVP--RPSQVSLKPAEEAHKE-----RKPKQPKXYTCQYCSRPCAKP 197
QY 121 SPLA-----TGO-----PAGESDLTKDRSLTGKLEPVSPSPPH-----TDPELEL 162
Db 198 SVLQKIRSHSGHEPYPYCGPFGSFKTSNLYKIR-----KSAHAIKGLASGSSSEM 251
QY 163 VPRILSKEELLQNDRDVREITWVEQOISLKKKKQOOLEEBAAPKPEKPVSPPIESK 222
Db 252 YPPGL-----EMERIPG-----EEFEPEPTEGSTSEBTGAASPGSTVLEPK---K 296
QY 223 HRSIVQIYD--ENRKAEAARLLEGLGPOV-----ELPLYNOSTRQYHEN 269
Db 297 HPLLSSILYSGSHSGSSQERCSLSQSGTSPLEDPAPFAASBHPLSHREDHTTKOK 356
QY 270 IKINQAMRKLLIFKKRNARQKQKPCQRYDQMLEALEKVERIENPRRARESKV 329
Db 357 LALNLSERKLLI--EQOTPLSPGSKSTSGTFSNSESABQOV---SPNTNAK--- 405
QY 330 REYYEKOPPEIRKQRELQERWQSGVQSGLSMAASRHEVESEITDGLSEQNLKOM 389
Db 406 -SYAEIIFKGC-----GRIGRTSMIASTQ----- 431
QY 390 QQLAVIPPMYADQOQIKFTINMGMLADPMKYKQKQVQVNMWSDEKETFREKFMQHPK 449
Db 432 ---PLPLSEDEPRLV--PLSVP-----RQVIEH-- 457
QY 450 NFGILASFLEKRTVAECVLYVYLTKNENYK---SLVRSYVRGKSGQOQOQOQOQOQ 505
Db 458 ---ITKLI---TINEAVVD---TSEIDSVKPRRSILTRSSVESPSS----- 496
QY 506 QOQOQPMRPSQOEKE 550
Db 497 ---LYRDLSSHGKTKQEOQLSLQHPSPSTHVPPLIRSHMPSAACTISTHHTF 550
QY 551 -DTSSEENDKEKAVASKGKTANSQGR---RKGRITRSMANZANSEATTPQOASLAS 606
Db 551 RGSYSFDDHVADEPVPBRNTPVFTSHRMLKRAAIELPLGGEYSSE---PPSSSDPT 607
QY 607 MELNESSRWTEEM-ETAKKGLLEHGNWGAIAWVGSKTIVSQCKNFYFYKKRQUND-- 663
Db 608 SKSDDEBPRESDLTKTKKGPFTKGANY-----ECTICGARYKKGRDYEAH 654
QY 664 -----EILQHKL--KMERK-----NARRKK----- 684
Db 655 KKYCYSELIQTKAHVGAHEVEKTOABEPWQMMHYKLGATLELPLRKRRKEKSLGDE 714
QY 685 -----KAPAAASEEAAPPVVEDEMEASGVSGNEEWEA 721
Db 715 EEPAPACPPSTAHNRPIGSTKSPAEASKSA---PSLEDPASSPGLPSQO----- 764
QY 722 EALHAGNEVPRGE-CSGPATV--NNSSDTESIPSPHTAAXOTQONGPAPPTLGADGP 778
Db 765 -----LQNGORRREOCPKKTIVIQHTSSPEKSP-----PEQSGLEEDKP 806
QY 779 P---PGPPTPPRTSAPLIEPTPASBATGAPPPPPAPSPSAPPPVVPKEKEEETAAAP 835
Db 807 PAQSSPPAPPHGRSAHSIQRLVRQ-----PNIQVEILVTEBDRPTEPEP 855
QY 836 PVEEGEOKPPAAEELAVDTGKAEVVSSECTEE--AEGCPAKGP--AEAAEATAEGA 890
Db 856 PPEPEK-----TEEFOWPQRTUQLAQLPAELPKPKKRLRLAEMAQSSGSS 903
QY 891 LKAKEKGGSGRATTAKSSGAPODSSATCSADVDVAEGDKURLSPR----- 942
Db 904 FESSVY--LNSPQOESSISLSSGSRASFPREDHGAE-----APGSPSDTRSKT 952
QY 943 ---SLTPTGPD---DANASPOK-----LDLKQIKQRAAIIPTIOV 978
Db 953 LGSMLTVPSSHHPHAREMRSASQSGPNVPHSSHMTETRSKSFYDGSLSPTGPSTL----- 1007

C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein

Query Match 3.78; Score 492.5; DB 2; Length 2774;
Best Local Similarity 20.5%; Pred. No. 6,4e-10;
Matches 553; Conservative 278; Mismatches 1002; Indels 861; Gaps 132;

```

Qy 77 ERSSEHLRPESHVYLPELGSSEMEFTESKRPRLLELPDLRLRSPILATQCPGSSDLT 136
D 77 ERSSEHLRPESHVYLPELGSSEMEFTESKRPRLLELPDLRLRSPILATQCPGSSDLT 136
D 398 DKAKKHLK-EKISKLEE-KDKEKKIKERKELKE-----EGREEK 440
Qy 137 KDRSLTGKLEVPSPSPHTPELELVPRLSK---EELIQMDVRDEITMVEQISKL 193
D 441 KDAKDEKRDKTEVKKLSKPDLPFTPEVKRLTYAKAPGRKVDK-----GRA 491
Qy 194 KKKQOOLEEBAKAPPEKPVSPPIESKRSIV-----QIYDENRKKAEMAHRIIE- 246
D 492 ARGEKLESEBPRTP-AQKGAAPPAVSGHRELALSSPEDLTQDFEELKREERGLAEQR 550
Qy 247 --GIGPOVELPLYNQPSD-TRQYHENTIKINQAMRKULILYFKRNHARKQKQKFCQRYD 303
D 551 DTGIG---EKPL---PADATEQGHPSAIIQVTPSGPVL-----584
Qy 304 QLMALKEKVRINNP-----RRAKESKYREYKOFPIRKQRELOERMQSRVGORG 358
D 585 --GHEVREKEVVDSPDCKGSTNRGPDSCAEVEKEKETWEERQREAE-----631
Qy 359 SGLSMSAAR--SEHEVSEIIDGLSEQENTLEKOMQOLAVIPMLYDADQOIRIKFINMGIM 416
D 632 LGPENTAARESEAEVEDVIEKALEBEMEE-----THPSD-----667
Qy 417 ADPKVKYKDRQVMMMSQEKETREKTMQHPKFGILASFLERKTYAECLVYYLYTKKN 476
D 668 -----EEGETAESEFYQKHQTEALAKASPSR-----694
Qy 477 ENYSKLVRSYRRRGKSQQOQQOQO-----OQOQQOQQOQPMRPSQOEKDEK 525
D 695 ---FALGRDGFQGGAKPEKETASPLSLATPAGATHEVSYIODETIPGSETQITISDE 751
Qy 526 EKEAEKEEE-KPEVENDKEDLLEKTDITDGEENDEKEAASKGRKTANSQGR---K 579
D 752 EIHDEPDERPAPRPFPTSTYDLSGREGPGPEASQADSAVPASSSKTYGAPETELTYPP 811
Qy 580 GRTRSMANEAENSEALITPOQSALASM-----ELNESPWTEEMETAKGLLEHGRN 633
D 812 NMVAAPLAEEHVASATSITECDKLSFATSVABQSVASLTAPQTEETGKSSLL-----866
Qy 634 WSAIARWVGSKTVSGCKNFYNYKKRONLDEILOHKLKMEKNNARAKKKKAPAAASE 693
D 867 LDITVTSIPSSRT-----EATQGLDIVPSAGIT-----SPTSSLEE 901
Qy 694 ---AAPPVVEDEMEASGVSGNEEVENVEAEALHASGNEVPRGCSGPATVNNSSDTE 750
D 902 DKGFKSP---CEDPSVTGSEK-----KGETVGRG-LSGEKAIVGKEEKV- 943
Qy 751 IFSHTTEAAKGTQNGKPRPATLGAQDP--PRGPTPRRTSKRAPIETTPASEATGATP 808
D 944 ---VTSEKLSGOYA---AVFGAPGHTLPPGEGALGEVERC-LISDDSDITVMASP-P 992
Qy 809 PRAPSPAPR---PVVPEKEKEE-----ETAAAPV--EEGEBQK--PPAABELAVD 854
D 993 PSGPSAAHTPFRHOSPVDEKSEPRDFODSWGETHSGVSKDESEEDTVKRGEEGISE 1052
Qy 855 TGKAEPEVKSCTEAEBSGAPKGAQDAEAATAE--GALKAEKKEGSGG-----R 902
D 1053 EGKG--PPTRSFOAQDMPVSIAGGCTGCTIQLPEQDKAIVETGBAGSNLAGTLPGEV 1111
Qy 903 ATTAKSSAPDSDSSATCSADVEDAEGCDKNLISRPBILITTPGPRANAPOKQLD 962
D 1112 TSTSEATEPQKDEVLRFTDQSLSPEDASLSLVASVSDPTT-----KQEAIPRSFCS 1163
Qy 963 LKQIKQRAAAIPIQVTKVHEPRREDAAPTKRAPAPRPPQNLQPSDAPQPGSSPRGX 1022

```

```

D 1164 LKEQPHKQILMPVSPEDTQSLSFSESPSEKET-SLDISKSQSLPSISGLTQEGELNLGK 1222
Qy 1023 -SRSPAPRPAKFAFAEAQQLPGDPPCWMSGLFPVPPREVRVIA-----PHADP-SAFS 1076
D 1223 EEEGPPVKAKE-----DSCCHLAIVSIPFHHRAVSPSTDETPAGTLPGSSFS 1269
Qy 1077 YA-----PGHPZPLGLH--DTARPVLPRPPTISNP-----PP 1107
D 1270 HSAISVDRKHSPEEITPGGHFMFTSDSSSLTKSPESLSSPAMEDLAVWBEKAPGKEKEPE 1229
Qy 1108 LISSAKHP--SVLEROIAGISQMSV-----QLHVPSSEKAP- 1144
D 1330 LKSETRQOKQOILPEKVAIVVEODLIHQKQALDEEKPFQODKTPEQKGRDLDEKDTA 1389
Qy 1145 ---VGPVTMGLPLPMDPKTLAPFSGVQKQQLSPRQ--APPPESLGVPTQAEASVLRGTA 1199
D 1390 AELDKRP-----EPKE-----KDLREDQGRAPPAE---KOKASEQDPTD 1428
Qy 1200 LGSVPGGSITKGIPTFRVPSDSAITYRGS-----ITHTGPADVLYKGTITRIIGED- 1250
D 1429 LQQT-----QATEPRDRAQERRDSEKQSLERARTBE--KQRI--LVQEDR 1473
Qy 1251 ---SPERLDR-----GREDSLPRGHVYI-----GKKGHVLSYEG--GM 1284
D 1474 APESHIPEPTQDRAPRDKGTDDKEQEBASEKEQVLEQKDWALCKEGEITLQOEARTA 1533
Qy 1285 SVTQCKSED---GRSSGPHNETAAPKRTY-----DMEGRVORALISSASIEGLMG 1332
D 1534 QKQETLEBEDTQOKSFPVEDKTTTSKETVLDOKSAEKASVQOQDAALAEKTRALL- 1591
Qy 1333 RAIPPE-----RSPHNLKEQHNIHG---SITQGIPRSYVEAOEDYLREAKLLKREGT 1383
D 1592 EESPAESKARQEKQKTKKQDVQGWRETSPIRGE--VGQKE-----PVPAWEGK 1642
Qy 1384 PPPPP---PSRDLT--EAUYKQALPKLKPRAHE--GLVATVKEAGRS--THEIPR 1430
D 1643 SPEQEVAVWRDRLITLQODAYWRELSCDRKWPFPHEIDGQAPRYCEBERESTFLDGP 1702
Qy 1431 EE---LRHTPEPLAP--RPLKEGSIITQGTPLKYDTGAGTSGKQDVASLIGSPERTF 1484
D 1703 EOEITPLQHTPRSGWTSDFDQFQEPPLQKGLVEERMLAESPVGLPPEEDKLTSPREIT 1762
Qy 1485 -PPVHPLDVAADARALERACYES-----LKSRPGTAS-----S 1517
D 1763 SPSPASPPRM--TGQRVSAQGESPVVDTESTAPMRKEPTTPSLALIPVWPKYDRLLP 1820
Qy 1518 SGGSIARGAPVIVBELKPRQSPULTY-----EDHGAPRAGHLPRGSPV 1560
D 1821 APLSPAPRPTAPPEPHTP--VPFSGWLAEDYSVVAVQGAABELGCPYS--PLGXDY 1875
Qy 1561 TMRPEPTRLQBSLSSSKAQODRKLTSTPREIA-KSPHSIVPPEHHPIPISEYELLRGVS 1619
D 1876 RKAEGEREGEGAGAPSSSPSPVPACGSLATRTDQTEPQREPTPYPDE---SFFQ 1932
Qy 1620 GVDLYRSHIPLAPDPTSPRGIPLDAAAYULPRHLA-----PNPTYPH 1663
D 1933 YADITBQMMTLGLCAPCTTEPPPLGASGM--PRLHSTKEBAAGCNTSAKETSSPASPO 1990
Qy 1664 ---LYPPYLIRGYPDTA-----ALENROTIINDVITTSQOMH 1696
D 1991 NLQSDTPAFSVASLAGAVRP---RQEPDPQPNVEPSTTPPAVPAPISLSKSLSPPLN 2047
Qy 1697 HNTATMAQR-----ADMLRGL-----SPRE- 1717
D 2048 GSTVCSGPRDRTPKETGKGHWDDGTNDDLEKGAAREKETSPPPHNPMWGHSSL 2107
Qy 1718 ---SSL-----ALNYAGPRGIIDLSQVPHLPVLVPPPTPGTATAM--- 1755
D 2108 WPTETAVSSLSDBHLSGVPSLDFPASAFGSSLQAP--POL--FSPAPRRAAPGSL 2163
Qy 1756 ---DR-LAYLPTAPQPPSSRHSS-----SPLSPQGP--THLYTPTT 1790
D 2164 AFSGDRALATVPGPHT--RTRHDEYLVETKAPSLDSSLPOLPSPSSPGGRLSNLPRPAS 2221

```

```

QY 1192 ASVLRG-----TALGSPVGSITK-----GI-----PSTRV----- 1217
Db 783 DSLAMGRREOQDITLSSDDFGSQLRHVDYEDSSSGSLPLPPOPPAADMTDEBFMRQ 842
QY 1218 -----PSDSAITRGSIHTGTPADVLYKGIITRIIGEDSPERLDGRGD-SLP 1264
Db 843 ILEMSAEEDNLEEDDTAVSGRLAGHGAQK-----ASARBPRESSQSVALP 889
QY 1265 KGHVITYEEKKH--VLSYEGMSVYQCKEDG----- 1294
Db 890 KRLPLHNAVTYIEELLSEEPAPFTDGLQGLRRFKTIGLSTGRMLSTLDLGQSGDP 949
QY 1295 -----SSSGP-----HETAPKRTYDMEGRVRAISASISIGLMRAIP 1337
Db 950 NLDBEPBELMESTLCSPEDRSGHSLPLPASTBYTGTSTLSLEED---SDSPS 1006
QY 1338 ERHSPHMLKEQ---HNRGSIITGIPRSYVEAOBYLRRKAKLLKRGSTPPPPSRDL 1393
Db 1007 RRRLLEAKQOKRKHRSHPPL--LPTIEDSEEBELREEBELLREQEKREVEVQQRIR 1063
QY 1394 TEAYKTQ-----ALGPLKKAHEGLVATVKEAGSHIEIRREELRTPPE 1438
Db 1064 STAKTRDKELRAQRERRSKTPPSNLSF-----IDBS-----PTEELRQAE 1109
QY 1439 LPLAPRLKEGSIITGTPDKTDTGASTGSKKHVRSLLGSPGTFPPVHPLDVWADARA 1498
Db 1110 M-----EELHSSCSGYSP-----SPSLDSEMETLDGGPTRLYKSGSEYNLPAM-- 1154
QY 1499 LERACYEESLKRPGTASSGGS1ARGAVIYVELG---KRPQSPULYEDHGAPPAHL 1555
Db 1155 -----SLCSPTETPGSGSTTPSGSRPLKSAEAYEDMMRKALLORQGOAAGARGP 1206
QY 1556 RGSPTWKEPRLPLOGSL-----SSKASDRKLTSPRRIASPHSTVEHHHPH 1606
Db 1207 HGGP---SQTPGPRQSGFEYQDTLDHDYGGASQP--AADGTPAGLG---ATV----- 1252
QY 1607 PISPYEHLRGVSGVDLYR--SHIPLAFDPTSIPRGIPLDAAAYLPRHLAPNTYPL 1664
Db 1253 -----YEELIQTOSIARMRQASSRDLAFTEDEKKKQOLMESAYMDMKONGSLTGT 1308
QY 1665 YPPYLIR--GYP-DTAALENQTIINDYITSQOMHNTATAMAORADMLRGLSPRESSLA 1721
Db 1309 SPTQLAAVSPFTSTSSDSSGGRVLPDVRTQ----- 1340
QY 1722 LNYAAGPRTIIDLQVPHLPVLVPT-----PGTATAMDLAYLPTAPQPPSSHS 1773
Db 1341 -HFAKEPQEPPLKHSPPASPLASKEVMTFSQOGTPTAT-----TMAPCPASLP 1390
QY 1774 SSPSPGPGTHLTKPTTSSSERERDRERDRERKSLITSTTVEHAPIMRPGTEQ 1833
Db 1391 RGYMTAPGER--SPSTSTHSHYQOPPTTANYGQTEB-----LPHAPSGPAGSGR 1440
QY 1834 SSSGSSGSGGGG-----GSSS--RPASHAHQHSF--ISPR-TODALQ 1873
Db 1441 ASREKPLSGGGEVGPQPSRGVSYFTGSSPPLSSTSESTSESTPLGLPRATAEFTQ 1500
QY 1874 RPSVLAHTMGKGIITAVEPSKPTVLKSTSTSPVAPPAITPPATCPGLGTYPLIM 1933
Db 1501 TPS-----LTPSS--DIPRSVGTSPVMAQGTQTP--HRPSTPRILVMOQSSQE 1544
QY 1934 EPVU---LPEKAPVAPRPERPRADT-----GHAF-LAKPPARSGL--EPASP 1975
Db 1545 APVMTITLASDASSQTRVTHASASTSPICSPDQSPASHSTSQTPPSPASQMPSEBAP 1604
QY 1976 SKGSEPRPLVPVSGHAT---ARTPAKULA-----PHASDPD-PAPPASADPHRE 2024
Db 1605 ---GPPRAPSAGVGPLALYGMGLPAMENISLCRISSVPGTSRVERGPRGTAVVDLT 1661
QY 2025 KTQSKXPSFI--QELFLRSLGYHSGSYSPGVEPVVSPSLTHDKGLPKHLELDKSH- 2081
Db 1662 AVKPTPIITLQGMULTSLAVEARKYG---LALDPVPGROSTAVQPVIVINLMQEQHT 1717
QY 2082 -----LBSELARPKQGPVKLGGEAHLPHLARLPRESQSSSPL----- 2119

```

```

Db 1718 FLATATTVSITMASSVMAQOKQPVVYGD-----PFGRLDFGQSGSGSPVCLAQKYVEQ 1772
QY 2120 -LQTAIP---GIVKHQRYVTLLAQH-----ISVITYODYRHHQQLASAPLPALY 2164
Db 1773 AVQTAIPYKGGPRGPRREKFAFYNYLPNQVTLARLDILITTMGAQVSLSKPGVPB- 1829
QY 2165 SFGPASCVDLRLRPPSDLYLPDPDHGAPARGSPHSEGGKSPPEPNKTSVYLGGE- 2219
Db 1830 -PGAE-----PHRATPALRLAHALPGTKPH---TVVQMEGGAAGT 1867
QY 2220 -----DGIEPVs-----PREGITE---P 2234
Db 1868 VTTLLPEEPAGALDLTGMRPEPSRLACDMAYKPPGSSCTGTFHPAPAPKSVTDALP 1927
QY 2235 GHSRSAYVPLLYRGEQTEP---SRMGSKSGQNTSQ-----PPAFSLUTEEN--SA 2281
Db 1928 GQSSGPPFYS--PRDPEPEPLTFPAQGVGPGPHEORPYDQGLPGRLYSSMSDNTLAE 1985
QY 2282 MYKSKQOEINKKMTNHN-----EPEYNI 2306
Db 1986 GLNTHAQRTIGQLFGQGRHDSAVDLSLKHYSYLGFPADGRYLQGLQYGSFTDLNHPD 2045
QY 2307 SQP-----GTEIFNMPAIT----- 2320
Db 2046 SHPLMPRVSSVSNISYDHRVGRPGDAVGFQEASLAQYSATTAREISMCAALNSMDQY 2105
QY 2321 -----GTGLMYRSCAVOEHA---STNGLBLIIRKALMGK--YQMEESPPLSANA 2367
Db 2106 GRHGGSGGPDLPVPQFQ---HGQGLNAPQGL--ASLRSGLLGNPTYPEGQPSFQNLAQ- 2159
QY 2368 FNPLNASAS-----LPAAMPITAADGREDHTLTSP----- 2397
Db 2160 YCPAASQGTAVRQLPSTATYRAADGMTYSTIINTPIAATLPTTOPASVLRPMRGMYR 2219
QY 2398 ---GGGKAKVSGRPSRKAKSP--APGLASG-----DRPVSVSVHSEGDQNRRT 2443
Db 2220 PYGSG---VTAVPLTSLTRPMIAPRVLPBAGLYRPAPSRPIASTI-----P 2267
QY 2444 PLTNRWEDRP-----SSAGSTPPP-----YNPLIMRLQAGVMASPPPG 2483
Db 2268 PABGPIVILGKPAAKASGAGGPPPELPAGAREBPLSTTAPPAVIKAPVAQAAPAPPG 2327
QY 2484 -LPA-----GSGPLAGP 2494
Db 2328 QKPADAAAGSGSVLGRP 2346

RESULT 15
A43359
microtubule-associated protein MAP1A - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Aug-1999
C:Accession: A43359, S22108
J:Langkof, A.; Hammarback, J.A.; Muller, R.; Vallee, R.B.; Garner, C.C.
J: Biol. Chem. 267, 16561-16566, 1992
A:Title: Microtubule-associated proteins 1A and LC2. Two proteins encoded in one messeng
A:Reference number: A43359; MUID:92355629; PMID:1379599
A:Accession: A43359
A:Molecule type: mRNA
A:Residues: 1-2774 <I>N>
A:Cross-references: GB:M83196; NID:g205537; PID:AAH48069.1; PID:g205538
A:Note: sequence extracted from NCBI Backbone (NCBIN:111039, NCBIPI:111040)
R:Citavchik, A.
submitted to the EMBL Data Library, June 1992
A:Accession: S22108
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 73-364, 'NLRs', 370, 'QKN', 374, 'PSFKGL', 381-751, 'RSMMSQMAQR', 764, 'D', 766, 'LR
'WIKRMGQPPQSP', 851, 'V', 853, 'NSL', 855, 'LPHRWLRN', 865, 'W', 867, 'HSQLEDGD', 877, 'Q', 879,
A:Cross-references: EMBL:X66840
A:Experimental source: strain Sprague Dawley

```

```

Db      1003 -----RCMLDKLRAFMERKPTI-----TACPTISKPDLVLYLYVYKRG 1045
Qy      1713 LSPRESSIALNYAGPRGI-----IDLSQVPHLPVLPVPTP 1748
Db      1046 FVEVTSKTKWMDLAGLIGLIGASSAAVTLRKHYTKNLLTFECHDRGDIDPLFIQV 1105
Qy      1749 GTPATMDRLALVLTAPQPPSSSRSSPLPGCGTHLTPTTSSSRERDRERDRDR 1808
Db      1106 GS-----KKTKAKASVP-SFGG-GHLDAGTNTGT----- 1134
Qy      1809 EREKSLITSTTVHAPIMRPGTESSGSSGSSGSSSRPASHAHQHSISPRTQ 1868
Db      1135 -----SSNSQDSFPA-PPGSAFNALIDGPGYPPGSPRGAS-----GPDIYAT 1178
Qy      1869 DALQAPSVLHN-TGMKGIITAVEPSKPTVLRSTSTSPVR-PAATFPATHCPLGTL 1925
Db      1179 AGQWRPSPQNNPQTPHPGAATAVAGD-----NISVNPREDIAG-----CGPG 1225
Qy      1926 DGYPTLMERYLLFKKAPRVARERPRADTGHAFAPASGLERASSPKSEPRPLV 1985
Db      1226 SGCTPG-----FGQGP-----GFGAASGA-----GAVGAVGGGPPHP 1259
Qy      1986 PPSVGHATIAITRA---KNLAPHASDPAPAPASADPHREKTSKPFSTIQLRLSLG 2042
Db      1260 PPRSPHTAAQAGCHQOQHONHONHONHONHONHONHONHONHONHONHONHONHON 1310
Qy      2043 YHSSYSPE---GVEPVSPVSPSLTHDKGLPKHLELDKSHLEGLRKPQGPVVLGG 2098
Db      1311 -GPPRAPQCHQGPQVPPSP-----QCHVRPAAGAPVPPGG 1345
Qy      2099 EAAHLPHLRPLESPQSSSPLLTAPGVKHQVYVTLAQHISEVITQDVTYHNOQLSAP 2158
Db      1346 SG-----YPTPVSRTPGSP-YPSQPGAYG-----QYSSDDYATGP 1381
Qy      2159 LPAPLSPGASCVLDRRPPSLYL-PP-PDHGAPARG-----PHSEGGKRS 2206
Db      1382 PGQPPGQGP-----QYPPQNNKMTFPGGEGAPPTANQYGPYSGRPYQPPGG 1433
Qy      2207 PEPNKTSLGSGEDIEFVSPPEGTEBHGSRSAVYPLLYRDBQTEBSRMGSKSPGNTS 2266
Db      1434 PQRPTQVAGGPRPAGAPGAPR-----SSAYP-----TGRPSQDQY 1470
Qy      2267 QPPAPFSLTESNAMYKSKQELINKLNTNKRBEPEINISQPTGELFNMALTG---TG 2323
Db      1471 QPPDOSQPPRRHDFID-----SOPYPGNA-----RQIYGAMQSG 1509
Qy      2324 LMTYRQAQVQENHSTNGLEALIRKALMKYDQWESPPSLANAFNPILNAGASLPPAMPI 2383
Db      1510 TQOYRPPQSSPAPON-----MGAPR---RGAAPRPGAHGPPIQP- 1549
Qy      2384 TAAAGRSHTLTLSPGGGK-----AKVSGRPSRKAQSPA-----P 2419
Db      1550 -AGVAMQMHVYPPQGGPRPQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 1608
Qy      2420 GLAGD-----RPPSVSVHSEGCNRTPLTNVWEDRPSASGTPP----- 2463
Db      1609 GQTAQSGIAPPGSLRPPS-----GPGQNNMPGMPAP--QOQSQOQGGVQPPQASH 1661
Qy      2464 ---TNPLIMRLQAGM-----ASPPP---GLPAGSGPLAGPHNADPEEKPL 2505
Db      1662 GGVSPFGILQVGGGMPVPPYAMPPPSGQGVQGVQGVQGVQGVQGVQGVQGVQGV 1712

```

RESULT 14

T42761

Bassoon protein - rat

N:Alternate names: brain-specific synapse-associated protein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T42761

R:Dieck, S.; Sammarti-Vila, L.; Langnese, K.; Richter, K.; Kindler, S.; Soyke, A.; Wex,

J. Cell Biol. 142, 499-509, 1998

A>Title: Bassoon, a novel zinc-finger CAG/glutamine-repeat protein selectively localized
 A:Reference number: Z22249; MIMD:98345363; PMID:9679147
 A:Accession: T42761
 A>Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-3938 <DIE>
 A:Cross-references: EMBL:Y16563; NID:93413503; PIN:CAA76287.1; PID:93413504
 A:Experimental source: strain Sprague Dawley; brain
 C:Function:
 A:Description: may be involved in cytomatrix organization at the site of neurotransmitter
 A>Note: component of the presynaptic cytoskeleton
 C:Keywords: coiled coil; zinc finger

Query Match 3.8%; Score 498; DB 2; Length 3938;

Best Local Similarity 19.5%; Pred. No. 6e-10;

Matches 506; Conservative 257; Mismatches 918; Indels 918; Gaps 114;

```

Qy      491 GKSGQGGGQGGGQGGGQGGGQGGGQGGGQGGGQGGGQGGGQGGGQGGGQGGGQGGG 550
Db      71 GVSRRIDPREKPLGSGQATSPTRKQASATAPGRSEPRP-----RA 111
Qy      551 DDTSGDNDEKAVASGKRTANSQGRKGRITRSMANENSEAATP-----QGSAL 604
Db      112 QGJSGQ-----EAGPRTLOVDSRTQSGR-----SPSVPPRGSTPTSPYSVQIAPL 161
Qy      605 ASMEINSSRWTEEMETAKKGLLEHGRNWSALABWVSXTVSGCKYFN---YKRON 661
Db      162 PSTLCPTCTSDLTSTSSQPNF-----NTCTCHANKVCNCGCFNPNH 205
Qy      662 LDEILQCHLMEKERNARRKKKAPAAASEAAPPVEDEMEASGVSGNEWEA 721
Db      206 LTVYKELCLNCQMQRALGMDMTTAPRSKSGQQLHSPALPAPSAKOPGKPEQ----- 260
Qy      722 EALHASGNEVPRGCGSPATVNNSSDTESITSP- HTBAADTGONGKRPATLGAADPP 780
Db      261 -----ERSRBPATQSGPRADEAARAT-SVPGTQATAPREVGVSPQPLSTKPTAEP 314
Qy      781 GPPTPRRTARAPLEP-----PASEATGAPT---PPA---PP 813
Db      315 RPPAGEAQSATTPVEGLAABEQGLTGKLFGLASLILTOASTLMSVQPEADTGGP 374
Qy      814 SPSAPPPVPEKEKEETAAAPVEEGEQKPPAAEELAVDTGKAEPEVSECTEAEG 873
Db      375 SPKSGPPKIYPSDASKAGRRP-----GSGGPP-----G 404
Qy      874 PAKGDAEAAEATAGA-LIAEKKEGS-----GRA-----TTASSGAPQDSBSAT 920
Db      405 PTPGAKTEPGRTPGSGPGALAKTGCTPSPKIGRADHQAASKAAAKPTMPKERACPL 464
Qy      921 CSADEVDEAGGDMN-----RL-----LSRPSLLPPT-----G 949
Db      465 QQA-ELNVGSRGPANNVNTTACKLRVCTLCGFNTPHIVKTEMLCLNCQTKLLBSLG 523
Qy      950 DPR--ANASPOKRLDKLQKORAAAPPIQVTKVHEPRREDAFTKAPAPAPPONLQ 1007
Db      524 EPAPLPLPTQEP--PAGVQRAAGASPLK---QKBPQGGQSGSLPRAAP-----QA 573
Qy      1008 ESDAPQGGSSPRK-----SRPAPRADKAPAAEAKLPGDPPTC----- 1050
Db      574 AKAPQAAKSPQAKPLRASEPSKTSASAEKKIGIVKAEPEVKKPPEETAVPGETPAK 633
Qy      1051 SGLFPVPREPVITASHPAR-----DPSAFSVAPRPHPLPLGADPT----- 1091
Db      634 SGVKTDPATPVPVAVDEAPKSGAEPEVPEKPYQSLSRFSQSLSDTGYSSDGYSSQSE 693
Qy      1092 -----ARVLPPTISNPPPLISSAKHPSVLERQIGAISQMSVOLHVPY 1137
Db      694 ITGVVQGEVQELDADAGVAGTPRPS-----PSELHKVGSMSRPLLEAQAAPAS 740
Qy      1138 SEHAKAPGVPTWGLPLPMDPKLAPSGYKQEQSLSRGQAGPPS-----LGVPTAOE 1191
Db      741 GEMSKPPSG-----SAVEDQKRPHSLSIMBEAFDSDELGLDLEED 782

```

Db 1465 HSNPA-----GIGQALQSLSSRGSGSVTAPGGHPR-HKCPGQAQOG 1504
Qy 1957 HAFKAPPAR-----SGLEPASPSKSGSEPRPVVPSYGHATTATFPAKNLAPHASNP 2009
Db 1505 PS--PRPTRYEPORVNSGL-----SSDPHFEEBGPMPVGVG-----TPRDSAG--VSP 1550
Qy 2010 DPAPAPASDPHREKTSKPSIOELEL-----RSLGYSGSYSPEGCVAPVSVPSLSLTH 2066
Db 1551 FPP-----KRRERPRKPELLOESLPPPHSGGFGSG--KPGC---PGFQASRDTG 1597
Qy 2057 DKGLPEKL-----EELDKSHLEGLRKPQGVKLGGEAALPHLEPLP--ESOPS 2115
Db 1598 TEALPHIMNLHTATSRKSYRPSMEPMMEPLSPEDVAGTEMGSQSDVLSGDSQVS 1657
Qy 2116 SSPLLQ-TAP--GVKGHQRVVTLAQHISEVITQDTRHHPOL---SAPLPA-PLYSFPG 2168
Db 1658 SGPGSGRSGSPDGLGK-----AABGPPKRGSSPLNAVPCGEPG 1698
Qy 2169 ASCPVLDRRPPSDLYLPPPHGAPARGPSHSEGGKSPENPKTISVLGGGDEGIEPVSP 2228
Db 1699 SEPP---RRPP-----PA---PHDGRKELP---REOPLPPGPIGTERSGRT 1736
Qy 2229 EGMTEPGHSRAVYPLVYRDGEQTEPSRMGSKSPGNTSQPPAFSGKLTESNSAMVSKKQ 2288
Db 1737 DRATEPGDIR-----PEHRGPPVQF----- 1757
Qy 2289 EINKKLTNHNREPEYINISQFTELFNNPALTGTGLMYRSGAOVEMASTMGLAIIK 2348
Db 1758 -----GTSDDXS-----DLRLVVGDSL---KABKE-----LTSYTE 1786
Qy 2349 ALMGKYDOWESSPLSANAFNPILNASASLPAAMPITADGSDHDLTSPGGGKAKVSGR 2408
Db 1787 ALPSPRD--WELLPSAASA--EPQSKNLDSGHCVPSPSSGQLYPEVYVGSAG----- 1837
Qy 2409 PSSRKAKSPAPGLASGDRPSPSVSVSHSGDCNRRTPLTNRWEDRPPSSAGSTP 2462
Db 1838 PSSSGISGSGHGLS-----ITSQWRRLRPPGRLPHFY 1869

RESULT 13
T13049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13049
R:Treisman, J.B.; Luk, A.; Rubin, G.M.; Heberlein, U.
Submitted to the EMBL Data Library, March 1998
A:Reference number: Z17592
A:Accession: T13049
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <TR>
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: FlyBase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the Wg pathway
C:Keywords: DNA binding

Query Match 3.8%; Score 502.5; DB 2; Length 2715;
Best Local Similarity 20.9%; Pred. No. 2.8e-10;
Matches 450; Conservative 189; Mismatches 752; Indels 763; Gaps 111;

Qy 670 KLNKEKERNARKKKKAPAA-----ASEBAFPVVEDEMEASGVSGNEEWEVEAEA 723
Db 4 KIKSPQTQQQQGGAPAPAAATPPSAGAAAGATPPT----- 39
Qy 724 LHASGNEVPGEGCGPATVNNSSDTESTIPS-----PHTBAKDTGQNGPKPATLGA 775
Db 40 -----SGPPTPNNSNNNGSDPSIOOOONVAFHRYGA-----PPPPGS--G 77

Qy 776 DGPPGP-----PTPRRTSR-----APTEPTPASEATGAP----- 806
Db 78 PGGPPGPRAAVMTYTHNLHQOOQHPPPHMQOOQHNGGRRAPPPGCAPEBAFGVKEYT 137
Qy 807 -TPPPAP-----PSPGAPPVVKEKEEETAAAPVEE---GEE 842
Db 138 HLPPPHPRAYGRVHADPNMDPRYGGDLPGGKPR---QQQOHPPQQPPQGPGGSP 193
Qy 843 QKPAEELAVDTGAKPEPVKSSETEAEBSR-----AKKDAAEAATLGLKAEKK 896
Db 194 NRPPQORYIPQPPQPGPTPLNLSLQSSNPPPPPHYANTYDQOAAASAAAAAQOQ 253
Qy 897 EGSGRATTAKSGAPQSDSATCSADVEDEAGCDKN-----RLSPR----- 941
Db 254 QAG--GPPPGHGPDPHQP-----PTGGQGGGAPPPRPSPQLGSGOQYRT 301
Qy 942 -PSLLPTGD--BRAN-----ASPQKELDKQKORAAAI-----PPIQVTK 980
Db 302 PPTNTSRGQSPYPYPAHQNGSYSSPQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 361
Qy 981 VHEPPREDAATKAPAP--PPQNL-----QPSDAPQPGSSSRGK 1022
Db 362 --QPQOQNTPTSGSYSPQRYPTPPLPAGSNHRTAYSTHOYPEBNRPMPGSSSPG 419
Qy 1023 SRSPAPADKFAFAEAKQKLPDPPCWTSGLPFVVPREVLIKASP--HAPDSAP--SYAP 1080
Db 420 SGHPLPPASPH--HVPPLQOQPPPHVSAGP--PP-----SSSPGAPSPSPSQASP 471
Qy 1081 GHPLPL-----GLND--TARPVLPPTISNPPLISSAKHPSVLERQIGALS 1126
Db 472 SPHQELIQNSNDSSGSAHSGMGSPGTGNTQVWRPFPSPGSS-----GSRG 522
Qy 1127 QGMSVOLHPVSEAK-----APV--GPVTMGLPLPMDKPKLAPFSGVQKQOQLSPGQA 1178
Db 523 MSPAVQNHPISRPASQSSSGGPMQOPVYAGGPPMPHPGWP--GAPPQOQOQOQOQ 581
Qy 1179 GPPESLGVPTAOEASVLRGTALGSPGSGITKGLPSTRVPSDASLITRGSLTHGTADV 1238
Db 582 SNSASSASNSPQQT-----PPRAPPPQGNMNAATPPPP--QGAAGGYPMPPH 629
Qy 1239 YKGTITRIIGEDSPRLDRGRBDSLPRGHVLYEKKGHVLSYEGMSVTQCSKEGRSS 1298
Db 630 MHGYSK--MGPPGQSPQAQGYPPQOQ-----QYRPNVPPRPPQYPPGAYAT 674
Qy 1299 GPPHETAAPKTYTMMGGRGRLASISASIGLMRAIPPE--KHSPPHL--KEQHIRG 1353
Db 675 GPP-----PPTSGAGAGANSMPSGAQAGYPRGCMHNTGOVPPYQWPPSPQOTVPG 729
Qy 1354 SITGIPRSYVEAQEDYLREBAKLLKREGTPPPPPSPRDLTEAVKTQALGPLKLPANEH 1413
Db 730 -----GAGCGAMVGNH-----VQKGTIPPPVVG-----GP-----PPQG 760
Qy 1414 LVATVKEAGSISHEIPREELRHTPELAPRLKESITGCTPLKYDTGASTGSKKHV 1473
Db 761 -----SSGPRPLNVL--KQHLQHKGYGSGSPTP-----PG--PGYGNQ--PTGM----- 800
Qy 1474 RSLIGSGRTPPPH-----PLDWAADARALRACYTEESLKRPPGTASSSG 1520
Db 801 -----HPGMPMGPRPHMGPPHTNMPPTSTPQOQMLQ-----GGQPOGQAGSG 847
Qy 1521 SIARGAVIPELQKPRQSPLYEDHGAPAGHLPRGSP-----VTWREP 1565
Db 848 -----PESGGRPH-----ISQDNQISSG--PTGAAGMAVTSVTTTGPDTGSDV 892
Qy 1566 TPRLQEGSLG--SKKASQDRKLTSTPREIASP-----HSTVPEHNPPI-----SPYENL 1614
Db 893 S---QOSTLNASASGSDPQCTTPKSKNDPYSQSHLAPRSTSPHVVHPPGGGPEEV 949
Qy 1615 LRG-----VSGVDLYRSHIPLAFDP--TSIRGCIPLDAAAYLYLPRHIANPTYPHLY 1665
Db 950 DMSGPPWMPRPASGPQVFNHVPOEFRTITTTTKKSDSLCKLY--EMDDNDR-- 1002
Qy 1666 PPLYIRGYPD--TAALNQRTIINDYITSQGMHNHTA--TAMQBRADMRL--G 1712

Db 5721 -----YEKK-----QAVHKEKVFIESFEEPYDELEVEYTEBPFEQPYEEBDE 5764
 Qy 263 TROHNENKINOAMKULI-----LYFKR---NHAKOMKOKFOR----- 301
 Db 5765 D---YEEIKVE---AKKEVHEMEEDFEEGOEYEREGYDEGEEMEEBAVOEREVIOVQ 5818
 Qy 302 ---YOUME-----ALEKK-----VERIENPRRRARAE---SKVREYEQ 336
 Db 5819 KEVVESEHERKVPAPVEKKAPPPPKVYIKKPIEIKETSRRAMEEKQVTKVPEVSKKI 5878
 Qy 337 FPEIRKQEOLEORM-----OSRVGSGSLGSL---MSARSEHEVS---EIIDGSEON 384
 Db 5879 VPQKSRTPVOEVEVLEVKVPAVHTKQWVISEEKMFASHTEEEVSVTVPEVQKEIVTEBK 5938
 Qy 385 LEKQKQOLAVIPMLYDADQ---ORIKFINNGLMADPMKY---KOROVMMMSOE 436
 Db 5939 IHVAVSKVEPPPKVPELPEKPAPEEVAVPVPI-----PKVPEPAKVPPEVPPKPVPEE 5992
 Qy 437 KETF---REKTMQHPKQNFGLIASFLERKTVAACVLYLTKNENYKSLVRSYRRBKS 493
 Db 5993 KKPVPVPPKKEPAAPK-----VPEVPKVPPEBKIPVPAKKKEAPPAKVPEV---QKGV 6045
 Qy 494 QOQOQ-----QOQOQ-----QOQOQOQOQPMRSSOE-----EKDEKEKEKE 528
 Db 6046 TEBKITTYTOREESPAPPAVPELPPKKVPEBKRPVPRKEEVPPEPPKVPALPPKKVPBEKV 6105
 Qy 529 A-----EKEEKEPEVENDKEDLKEKTDGSDGDNDEKAVASGRKRTANSQGRKKRIT 583
 Db 6106 AVPVPAKKAAPPRAPEVSKKTVV-----BEKRFVAEEKLSFAVPQ---RVEVT 6150
 Qy 584 RSMANANSEEAITPQOSAEIASMEL-NESSRMTSEEMETAKKGLLEHRNM-----SA 636
 Db 6151 R---HEVSAEEMWSISEEBEGVSIISVYEEEREEBEEAEVTEYEWEEPEEYVVEBKLIH 6207
 Qy 637 IARWVSKR---TVSQCKNFYFNKKRONLDE-----ILOQHKLMM---EKER 677
 Db 6208 ISKREAPAPAEVTEBOEKKIVLKPKIPAKIEBPAPKVEAPKKIIVPEKVPAPVPEKKEK 6267
 Qy 678 NARKKKKAPAAASGEAAPP-VEDEMEASGVSGNEEVEEALHAASGNEVPRGEC 736
 Db 6368 VPPVPEPEPKKPVPEKKVPKPKVIMKEBPPLAKVTEKIMQITOEKVLVA---VTKKEA 6323
 Qy 737 SGPAIVNNSDTESTISPHTEAQKDTGONGPKPPATLG-----ADGPPPPPT 784
 Db 6324 PPKARV---PEEPKRAVEEKKVLKIKKRESEPPAKVTEFRKKVYKKEKSVISAPKPEQ 6380
 Qy 785 PPRRTS---RA-----PIEPT----- 797
 Db 6381 PIKEVTIMEKERAVTLEEBAVSVQREBEVEYEVEYDYKPEPEYEPTEEDQVEYEBRE 6440
 Qy 798 -----PASEATGAPT-PPAPPSAPPPVPVPEKEKEEETAAA-----P 835
 Db 6441 YERYEEHEEYITEPEKPIPVKPVPEEPVPTKAPPAKVLKAVPEEVPVPVPIPKKPP 6500
 Qy 836 PVEGEEOQPPAAAEIADVTKAEBPVSECTBEABEGPAKGDABAATAAGALXAK 895
 Db 6501 PPKVPEEKVYFEEKIHISTKRE---KEQVTEPAKVPMPKPK-----RVAAEEKVPVPR 6552
 Qy 896 KEGSGGATTAKSSGAPODSDSATCSADE---VDEAEGDKNRLSPRSLTJ----- 946
 Db 6553 KE-----VAPRVAVEVPKELEPEEVAFAFEVEVTVHVEEVLVEEBEYIHEEBEFTTEEBV 6608
 Qy 947 ---PTGPRANAASF-----QKPLDKQLKQKRAAATPPIQVTVHPRPRDAPPTPAPAP 999
 Db 6609 PVIVKVPVEVPKVPPEKKVPVPPKKG---APPAKVPPEVPKPKPEKVPVLPPKKEP 6664
 Qy 1000 PPOMLQESDAPQOOGSSPRGKSRPAP-PADKBAFAAAEQKLPGDPPCMTSGLPVPV 1058
 Db 6665 PP-----AKVPEVP-KKVPPEKVPVVPVKKYEAAPPAKVPPEVPPK-----PVP 6706
 Qy 1059 PREVIKASPHAPDDSAFVAPPGH-PLPLGLHDTARVULPRPPTISNPPELLISSAKHPS 1116

Db 6707 EKVVPVPAKVE-----APPAKVPPEVKKLIPEBKKTPTVPKKAPEAPPPKKEBP 6759
 Qy 1117 VLEROIGALISQMSVQVHVPVISEHA--KAPGVPTMGIP-----LPMDPKGLAPSGV-- 1167
 Db 6760 -----VPPVPAVPOEEVULPEEELVPEEVLPEEVEVLPEEVEVLPEEVEVL 6807
 Qy 1168 KOEOLSPKQAGPPEPSLGVPTAOEASVLRGALGVSFGSITTKGI-PESTRVPS-----DS 1221
 Db 6808 EEBELPPEEVEVPPEEVEVPEEVE-----FVEEVEVLPEKVPVPAVPAPELKK 6857
 Qy 1222 AITYGSI-----THGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEGKKGHV 1277
 Db 6858 KYTEKVVAPIKKEAPPAKV-----PEVPKVEEKA-----IILPKEBEV 6897
 Qy 1278 LSYEGMSVYTOCSKEDGSSGCPHETAPAPKRTYDMMEGRGRAISSASIGLGRAIP 1337
 Db 6898 LPVE-----VTEPEPEEPISEBEIPEEPPSIEVEEVAAPRVPAVYIKAVPBA--PTVPK 6951
 Qy 1338 ERHSPPHLKEQHNRGSIITOGIPRSVYEAQEDVLRREAKLLKRGSTPPPPPSDLEAY 1397
 Db 6952 KVEAP-----PAKYSKIPBEKVPVPO---KKEAPPAKVPPEVPPKVPBEKVLVP-- 6998
 Qy 1398 KTOALGPLKLKPAHGLVA-----TVKE-----AGRSIHEIPREELRTPELPLAPR 1444
 Db 6999 KKEAVPPAKGRTVLEBKVSVAFRQEVVKELELEVEAEVBEIPEEBEFHEVE-----E 7053
 Qy 1445 PLKGSITIGTPLKYDTGASTGSKGDVNSLJSSPGRTPPVPLD---VMADARALE 1500
 Db 7054 YFEEGERF-----HEVEEFILQHRVEEHRVEKVNRIIVFEAE 7094
 Qy 1501 RACEESLKSRRPGTASSGGSIGANGAP--VVPPELGKPROG-----LTVEDGAPFAG 1552
 Db 7095 VEVEKRP-KAPPKPEISEKILPPKKPTPKVPPEKKEPAKVPPEKPIVBEKVRP-- 7150
 Qy 1553 HLPKGSVPVMEPTPRLQOGLSSSSKASQDKLSTPBEIAKSHSVTVENHPHPSYE 1612
 Db 7151 BEPVPPTKVPEVLPKPEVPEKVPVPAKPPAPPPKVEAPKVEPEKVV-FVPPPK 7209
 Qy 1613 HILRGVGVULYRHHIPLAPRPTSIPIRGIPLDAAAAYLPHILANPPTYPLPYLIRG 1672
 Db 7210 -----KEVP---PTKVPF-VPPAAVPEKKVPALPPKESP---PEVFE 7248
 Qy 1673 YPDTALENRQTIINDYITSQMHNTATAMAQADMLRGLSPRESSLA-----LN 1723
 Db 7249 ----- 7267
 Qy 1724 YAGPRGIIIDLSQVPH-----LPVLVP-----PTGPTATAMDRLAYL 1761
 Db 7268 VPEVPKEVPEKVPAPAPPKKPEVTPVKVPEAPKVEVPPEKKVPVPPKKEVPPTKVPEV 7327
 Qy 1762 PTAQPPSSRHSSSPPLSPGPTHLTKPTTSSSRERDRDRERDREREKSLITSTTV 1821
 Db 7328 PKAVPEKKVPEALPPKPEEP---PEVEPEEPEVALPEBPAAVEBPAPAPQVTV 7383
 Qy 1822 EHAPIMPTGEOSSGSSGSSGSSSPASHAHQHSPIPSRTODALQOORSVLAHT 1881
 Db 7384 PKNV-----PEKKAIVAAKKRELPV-----KPEVPEK 7415
 Qy 1882 GMKGIITAVEPSKP-----TVLRSTSSVPRPAATPPATPCPLGGLDGVY 1929
 Db 7416 VPEKKVPLVVPKPEADPAKVPPEVPEKKAIVPKPEV---VP 7463
 Qy 1930 PTLMEPVLLKKEAPRVARPPRPADTGHAFIAPRPAASGLEPASPSEK-----GSPRP 1993
 Db 7464 EVPPKVPV---BEKPAVPEPRAASPPPEV---BEPEEIAEPEEIAPEEKVPVPAEPEEP 7519
 Qy 1984 LVPVSGHATIAIPAPNL-----APHHAAPD-----PAPPASDP 2021
 Db 7520 EVPP-----PAVEBPKKIIEKVPVYIKKEAPPPKPEPEEKVIEKKLKRPPPPPPAP 7575
 Qy 2022 HREKTOGKPSI-----QELLSRLGYHGS-----SYSPEGVAPVS 2057
 Db 7576 PREDEVKEKIFOLKAIIPKKVBPENQVPEKVELPPLKVPGGKKVKKLLPERKPEPEEVP 7635

```

QY 1339 -RHSNHLKEQHHI-----RGSITQGIPIRSYVEAOE-DYLREAKL----- 1378
Db 1855 RSHAPRGKRHHYVWQMBGCTAGTVTLLEBPAGALDLGMRPEGLACDMVYKPF 1914
QY 1379 --KKEGT--PPPPPSRDUTE-AVKTQALGPLKKAHEGLAVTAKGASIHIEI PBEEL 1433
Db 1915 GSSCTGFHNPASAPADKSVTDJALPGOSSGPF----- 1946
QY 1434 RHTBELPLAPPL-----KEGSITQGTPLKY-----DTGASTTSGKHDPV 1474
Db 1947 -YSRDEDEPPPLTFRTQGVVGPGRHEQRPYQGLPERLYSMSDTNLAAGANTYAKR 2005
QY 1475 --SLIGSGRTFPPVHPLDVADARALERACYESLSKSPGTASSG--GS---IARGA 1526
Db 2006 LGQLFGQGR-----DSAVDLSLK--HYSISLGADGRYIGQLQYQTSFDDLRPT 2054
QY 1527 PVIVBELKPRQSPPLT--YEDHGAPFAGHLPRGSPVTMRPEPTPLQEGSLSSXAKQDRK 1584
Db 2055 DLISHPLPLRKYSVSNIYSDH-----RYGRGDAV-----GFQESLQYSA----- 2097
QY 1585 LTSTPREIAK--SPHSTVPEHHHPISPYEHLGLGVSVDLYRSHIPLAFPTSTIRGI- 1641
Db 2098 --TTRAEISRMCAALNSMDQYGRHGS-----GSGGPDLYQ-YQPGHGLSAPQGLA 2147
QY 1642 PLDAAAAYLPRHLAPNPTYPHLVPRYLIRGYPTALENRQTIINDYITSGQMHNTAT 1701
Db 2148 PLRSG-----LLGNPTP-----EGQPSGNLAQYGRASQATAVRQLLPSTAT 2191
QY 1702 AMA-----QRADMLRGLSPRESSLALNTYAAGRGIIDLSQVPHL 1740
Db 2192 VRADGMISTINTPIATLPITTPQASVLRPMV--RGMYRPVYSGGVAVPLTSLTRV 2249
QY 1741 PVLVPTPTGTATMDRLAYIPTAPQPSRHSPLSPGPTHLTKPTTSSSEREDR 1800
Db 2250 PMIPRVPLGAG-----LYRYPAPRPPIASSVPPAE-GPVYLGKPAATKAGAGGRP 2301
QY 1801 DREDRDRERKESI-LTSTTVEHAPIV-----RPCTEBSGSGSGSGGSSSPRA 1852
Db 2302 RPELPAGVAREEPSTTRPAVKEAPVAPABGAPAPRPQKPRAGEAAGSGSVLSRPA 2361
QY 1853 SH----- 1854
Db 2362 SEKEASQEDRQKQEQQLQLEBERVELEKLRQLQLEBERERVELQRHEEQQLVQ 2421
QY 1855 -----SHAQHSIPRTQDALQORPSVLAHNTKGI-----TAVBPSK-- 1894
Db 2422 RELQELQIKQHVLCQQQEEEROAQFALQREQLAQRLQLEQIQLOQLOLEBOQKORQ 2481
QY 1895 -----PTVLRSTSTSPVPAAT-----PPATHC-----PLGGTLGCVPTLMEPVV- 1937
Db 2482 KAPRPATCEAPSRGP-PAATELAQNGQVMPPLTHAFAIYAGTBGQGP--REPVLAHR 2537
QY 1938 -LPEAPRVARPERPRADTGAFLAK--PPASGLEPASSPSKGS----- 1979
Db 2538 GLPESASDMSLQTEBQWAGRSIGIKKRHSWRLRACGSPSGPDSYTRIAIDSSVQDD 2597
QY 1980 --EPRPLV-----PVUSGHATIAETPAKNAIAPHASDP 2011
Db 2598 EEEGRGRVYTRRRRTRSADCSVQTDDBDNADWEQPV-----RRRSRLSRHSDGSD 2650
QY 2012 PAPASADDPHREKTOQSKPFSIOLELBSLGVHSGSVSPEGVEPSP----- 2058
Db 2651 SKHDATASSSTTAATAAPAMS--SVGIOTIDCSVQTEPEQIPRVSPAIHTIATDPRVE 2708
QY 2059 -----VSSPLTHDKGLPKHLE--ELDKSHLEGELR 2087
Db 2709 IVRYIASAEKTOGRGSLACQTEPDQAGVAGPQLIGFTALSPVLRGQIYVPGALGRFE 2768
QY 2088 PKQCGPVLGGEAAHLR--HLRPLRSPSSPPLQTPKVGKHQRVYVTLAQH----- 2138
Db 2769 KKKDPLEIIGYQ-AHLRPEESLSQVSRQPKSPQVLYSP-----VSLPLPHRLLDTSF 2820
QY 2139 -ISEVITQDYYRHHQQLS-----APLPALYSPFASCPVLDLRPPSLYLPP-- 2188

```

```

Db 2821 ASSERLKAHVSPQKQFIADSTLRQOTLPRPMKT-----LQRLSD---PKELSP 2867
QY 2189 --DHGAPARGPSHS-BEGKSPENKTSVUCGEDGIE-----PVSPEGWEP--GH 2236
Db 2868 TAEESAKERESLYOHQGLGS---QVSVL--PPNGLVRYKRTLPSPPEBAHLPLAQ 2921
QY 2237 SRSAYV--PLLYRDGEQTEPSRMGSKSPGNTSQPPAPFSKLTESNAMYKSKOEINKL 2294
Db 2922 VPSQLYAASLLQR-----GLAGPTVPATKASLLELDRLALVEHESTKAKKQ 2971
QY 2295 NTHNRNPEYNIQOPGTEI-----FNNPAITGTGLMTYRSQAVGEHAS-- 2337
Db 2972 AEIDEEKEIDAKLYBELGITQKESLAKORGRDVPPLRGIG-----EHRDYL 3021
QY 2338 -----TNMGLEAIRKALMGKYDOMESPPLSANAFNP--LNASASLPAAMPITLADGRS 2390
Db 3022 SDSELNQLRLQCTTPA--GQYVDYPAASAIVPAPPSGPTAFQOPRFPAPQYTAGS-- 3076
QY 2391 DHTLSPGGG---KAKVSGRPSPRSKAKSPAPGLAGDGP--PSVASVSEGDGCRNRT 2444
Db 3077 ---SGFTQNGPRAHQAPTYTGESTYPAPTYPTGTGPAEBGLPSQAFHPTGYAAPT 3132
QY 2445 L-----TNRYWEDRPSAG---STPPYNPILMLRQAGVWASBPPEGLPAGS 2488
Db 3133 MPTQSAFPYQVADSRRAHQKPRQTSLADLEQKVPYTVYVIGSPAVTMSAPFE--TGYS 3190
QY 2489 GPLAGPHHAWDEPKPLCSQYETLSDS 2516
Db 3191 GPVAGSGYEOGKAPENHPRGSDRSVSQS 3218

RESULT 11
138346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: 138346
R:Label: S., Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: 138346
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 3.8%; Score 503.5; DA 2; Length 7962;
Best Local Similarity 19.0%; Pred. No. 8.2e-10;
Matches 534; Conservative 355; Mismatches 1053; Indels 865; Gaps 128;

QY 27 PVQIARTHTDVGLLEYOHSHRDVASHLSPGSIQPPORRRBSLSEFPQGNERS-----Q 80
Db 5498 PIQFTKIQNVASE--HQSATFECEVSFDADATYWKGPTELTESQKYFRNRDGRCHYM 5555
QY 81 ELH-LRPSHSYLBELKSEME-----FIESKPRLEL-----LPD----- 115
Db 5556 TIHNVTPDDEGVSVIARLEPRGARSTABEILYLTTKGIELEKLPDPIDPSVPIPTMIR 5615
QY 116 -----PLLRBP-----LLATQGPAGSED-----LTQKSLTGKLEPV 148
Db 5616 AVPEEIIIPVVAAPVPILLPTPEBEKPPPKRIEYTKKAVKADAKVAVAKPEMTPREBIV 5675
QY 149 SPSPRHTPDBLELVPRRLSKELIOMMDRVDRITMVEQIISKL-KKKQOQLEEEAKP 207
Db 5676 KRPPEPTT-----LIPAK--APEIIVDSKAE-----EYKIMTKRKEQVKEKEAV-- 5720
QY 208 PEPEKPVSPPIESKHSVLQIYDENRKKAEAAHRIELGIGPQ-----VELPLYNOPSD 262

```

QY	2338	TNMGLEAIRKALMGKYDQWEESPPLSNNAFNPLNANASLPAAMPITRAADRSHTLTSP	2337
Db	1782	----LTASVTAIPAIVSRD--WELTSPASA-A-EPOSKLDSGCHVPEPSSSGQRLLPEVFY	1835
QY	2398	GGGGKAKVSGRPSRKKKSPAPGLASGDRPPSVSVHSEGDCKNRTPLTNKRWEDRPSA	2457
Db	1836	GSAG-----PSSSQISGGSHGLS-----ITSKQMLRPGTTP	1866
QY	2458	GSTPF	2462
Db	1867	SLHPY	1871

RESULT 10

T42730
Bassoon protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42730
R:Deek, S.; Sammarti-Vila, L.; Langnaese, K.; Richter, K.; Kindler, S.; Soyke, A.; Wex, J.; Cell Biol. 142, 499-509, 1998
A:Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively localized
A:Reference number: 222249; PMID:98345363; PMID:9679147
A:Accession: T42730
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-3942 <DIE>
A:Cross-references: EMBL:Y17034; NID:G3413809; PIDN:CAA76598.1; PID:G3413810
A:Experimental source: strain 129 SVJ
C:Genetics:
A:Map position: 9F1
A:Introns: 72/2; 208/3; 505/3; 675/3; 2889/3; 3582/1; 3851/3; 3886/1; 3930/1
C:Function:
A:Description: may be involved in cytomatrix organization at the site of neurotransmitter
A:Note: Component of the presynaptic cytoskeleton
A:Keywords: coiled coil; zinc finger

Query Match	3.8%	Score 504;	DB 2;	Length 3942;
Best Local Similarly	19.1%;	Pred. No. 3.7e-10;		
Matches 590;	Conservative 324;	Mismatches 1122;	Indels 1052;	Gaps 136;

```

QY 55 PGSIIOPORRRPPLSLSEFOPONERSQELIHLRPESHSLVLPGLGSKEMFIESKDPRIELLP 114
Db 557 POGIGOGSSGSLPALAS-----PQATKASPOATKASPOATKASPOTTKASP 601
QY 115 DPLIRPSPPLATGOGPAGESDITKORSULT-GKLEFVSPSPPHDTPELELVPRLSKEELT 173
Db 602 ---QAKPLPATERPSKTSSTSSAOEKKTATPAKAEV--EKPEPET-----TVPCGPYAK-- 649
QY 174 ONMDRVAREITMVEQOISKLKKKQOOLEEFAKRPPEKVPSPPEPTESKHSLSVLQIUYDE 233
Db 650 SGVKTDPPIAPVVKP-----VPEAPKGEAEVEVPKYSODLSRS----- 689
QY 234 NRKKAEAHRIEGLGPOVELPLYNQPSDTRQCHENIKINOAKRKLIIFYKARNHARKO 293
Db 690 -----PQ-----SLSDTCYSSDGGVSSQS----- 708
QY 294 WKQKFCQRYDOLMEALEKVEYRIEN--NPRRAKESKVABEYKQPEIRKQREIQERM 350
Db 709 -----EITGVQGEVEGQLDSAGVTPRPPSPBELIKVSSLSRPSLEAQVAPSAE 758
QY 351 QSRVGQRGSGLSMSAARSEHEVSEIITDGLSEQENLEKOMQOLAVIPMLVYADQORIKFT 410
Db 759 WSKRPRESSSAVEDQKRPHSLS-----IMPEAFDSDEBEGDIL 797
QY 411 NNMGLMADPMKIVYKDRQVMNMWMSOEKETPREKFMQHPKNGFLIASPLEKTYAECVLVY 470
Db 798 EEDDSLAE-----WGRQR-----EQQDTAE----- 816
QY 471 YLTKNENYSYLVRSTY---RRRGKS---QOQOQOQOQOQOQOQOQPMRPSQOEKUE 522
Db 817 ---SSDDFOSQORHDDVEDDSSEGLSPLRPQPPARADMTDEPFMRQITLMEABEDNUE 872

```

QY	523	KE-----	KEKAEKEEKPEVENDKELLKEDTDTGENDUDEKAAVASKG-RKTA	572	
Db	873	EDTAVSGRGAKISAOQASAPRPESSOEPRRRIPPHATTGYE----	ELISEAGPAIPT	928	
QY	573	NSOGRKRRIRTRSMANENSEEAITPQOASAEIASMEINES-----	SPWTEEME--TAKK	625	
Db	929	D6GALOGGLMRFKTIELNS-----TGSYGHEL-----	DLQGDHPDMLDRPELEMESLTGSP	981	
QY	626	GLIEGRMWSAIAHWGSKTVSOCKNPFYNYK-----	KRONDELLOOHK-----	670	
Db	982	EDRSGEHSSTLPASTPYSTGTSPTSLISEEDSDSPSRORLEEKQOKRNRSHG		1041	
QY	671	-----LKMEKE-----	RNARKKKGAAMAASEAFA	696	
Db	1042	PLUPTIEDSSHEBELREBELLEDEOKKREVEQORIRSTAKTRDKNELAQAORRREKSK		1101	
QY	697	PP-----VDEEMEGASGVSGNEEMWEABALHAGNEVEBRCGCPATVNNSSDTESI		751	
Db	1102	TPPSNLSPIED-----ASPTIELRQAEM-----	ELHRSSCS-----EYS	1137	
QY	752	PSPTHEAKKQNGQNP-----	KRALTLGADGP-----	PPGPTPPRTSRAPIEPT	798
Db	1138	PSPSLDSAEKTLUDGGPTRLKYSGEYMLPAWSLYSPETTSGSGSTTP--	SSGAPLKSAE	1195	
QY	799	AS-----	EATGATPTPPAPSPAPRPVNVPEKEKEEETAAPAVEEGEE	842	
Db	1196	EAYEDMRKKAEMLOQOQOVAGAGPHGCPQPTG-----	PRQSGPFYQDTORHNDYGR	1256	
QY	843	QKPPAEELAVDTGKA--EEPVKSECT-----	EEABEGPAGKGDMAEATLEGALKA	893	
Db	1251	ASOPAEESTPAGLGAAYVEEILLOTSOSIARMRQASSRDLGPTEDKKKEKOFLENAESAAMD		1310	
QY	894	EKKEGSGS-----	GRATYAKSGA-----	POD-----S	915
Db	1311	PMKONGGLTPTGTSPTQOLAAPVSVSTSTSSSSGGRVAPDVRYQHPAKHEQODLKLHSS		1370	
QY	916	DSSATCSADEVDE--AEG-GDKNLLSPRPSLLTPTGDPFRANASPPQRLDKOLKORAA		972	
Db	1371	PVSSTLTSKEGGMTFSGQGPATTAATPRGYMPTPSGERSBPSSTHSYGQ----		1426	
QY	973	IPPIQVTKVHEPRREDAAFTKPAEPAPP--PQONIQPSDPAQOQSGSPRGKSRAPAPAD		1031	
Db	1427	-PPTTANY-----	GSQTEELHAPASGPPGSGRAPAREKXLSGDSDEVG--	APQBSRG	1474
QY	1032	KEAPAAEAKOXL-----	PGDPFCMTGLFPFVPRREVIAASFPAPBPASAFSYAP--	FGHPLP	1085
Db	1475	YSYFTGSSPLSPSTPESBSPFSG--	KLGPRATAFSTQTSPBLTSLSDIPRSGPPSP	1531	
QY	1086	LGLHDYARFVLPREPTI-----	SNPPLISAKHPSVLEROIIGAIISQGMVSQVLHVPSER	1140	
Db	1532	MVAOGTQTPHAPSTPRPLWQOSSOEAEIM-----	VITLADSDASQOTRWV-----	H	1576
QY	1141	AKAVGVV-----	TWGLPLPMDPKXLAAPSGVQBELSRGQAGPESLIGVPAQASVLRG	1197	
Db	1577	ASASTSPLCPTDSQPTSHYSIQTTPPSAGOM--	DSEBGRGP--	GFPAPASGTDGP	1629
QY	1198	TAL-----	GSVPGSGITTKGIPSTRVPSDAITRYSI		1229
Db	1630	LALGWMGLPAENISLCRISVPPTSRVVE--	PGBRPGTAVVDRLTAVKPPTLITLQGM	1687	
QY	1230	-----	TH-----	GTADVLVKGITTR	1245
Db	1688	DLTSLAVERKYGALDPVSGROSTAVOPLVINIMAOEQHTFLATATTVSITMASSVLM		1747	
QY	1246	-----	IIGEDSPSLDRGREDSLFP--	KGHVITYEGKKGHVLSYEGMSVTOQSKEDGRS	1296
Db	1748	AQOQAPVYVGDPPGSRULDPQOGSSSPVCLAQVQVEBAGVOTAPYRGG-----		1794	
QY	1297	SGGPHET-----	AAPKRTYDMMEGRVGRALISSASIEGIM-----	GRALPPE--	1338
Db	1795	PRGPRFRAKFARYVLPNOVUPLARBDILITOMGTAQGVGKPGVBPVBPGEAPHATPAEL		1854	


```

Db      2352 LVKTTKSPAEWTPRRIDKLKEMVPHSEYEAFAVPAHGEODES--GSPNNMFLQPLD 2410
QY      943 SLTPTGPRRA-----NASPO-----KPLD---LQKQRAAAIPIQYT 979
Db      2411 QELTSSQPHGMVPHNPTPGKIYLHYAEPPTGPFVEPRDLFLTKTKSKPVQSGPKELA 2470
QY      980 K-----VHSPREDA-----PTKPARPAPPP-----ONTLESDAPOQPGSS 1018
Db      2471 KSPKEMVQTPREYKAAVLSCPEODESPSPNNMLKSLDOEVMSSQPHSGVPIRP--KT 2529
QY      1019 PRGK---SRSPAP-----PADKEAPAAEQCLPDDPCWTSIGLPPVPVPREVIK---A 1065
Db      2530 P-GKIYLHSIEPRPPPFVKEPTDLILVTKTKSKPAB--WT-----PRRIDKLKEM 2577
QY      1066 SPHADPFAFSYAPRGHPLPLGLHDTARVLPRLPRTISNPRPLISSAKHPSYLEQIGAI 1125
Db      2578 VPHSE-----VEBAVFP-----AHGEQDESGSPNNMFLQ-----LDDELTLIS 2617
QY      1126 SQGMSVOLHVP-----YSENAKAPVGPV-----TMGLPLPM--DPKCL 1161
Db      2618 SQPHGMVPHNPTPGKIYLHYAEPPTGPFVEPRDLFLTKTKSKPVQSGPTQOMAKSPBEM 2677
QY      1162 APFSGVQEOULSPRQOAPRPESLVGVPTAOEASVLGTAAGSVPGSITKGIPISTVPSS 1221
Db      2678 VSLSPKNEVEVFPAAQCGODESPSPNNISLOLDOEITMSQPHGMIPH-----PPN- 2729
QY      1222 AITRAGSITHG-----TRADVLYKGTITRIIGEDPSRLDGRDLSLKG 1266
Db      2730 -----THGKIYLHYAEPPTGPFVEPRDLFLTKTKSKPVQSGPRQIDSKPKMTQO- 2780
QY      1267 HVIYEGKGAHVLSEYCGMSVTQCSKEDRSSGPHETAAPKRTYDMEGRGVRAISSAS 1326
Db      2781 -----SPYEESLPLRAHAEQGEESRAPRHFSLQF-----LDQSLSSLS 2818
QY      1337 IEGLMGRAIPRRHSRPHLKKEOHNRIGSIITGIPRSYVEAOEDYLARRAKLKEG---- 1382
Db      2819 -----HPHGMIPHNPTPGKIYLHYAEPPTGPFVEPRDLFLTKTKSKPVQSGTATK 2869
QY      1383 TPRPPSRDLTEAKTQAL--GP-----LKLKRAHEGVAIVKAGRSIH--E 1427
Db      2870 TDKSDEDEVSGQPREYKAVALSGPGEODESPSPNNMSLOSLDOELAISSQPHGMIPHPN 2929
QY      1428 IPREELRTPELPLAP-----RPLKEGSIQO-----GTPLKYDTG 1462
Db      2930 APDKIYLHYAEPPTGPFVEPRDLFLTKTKSKPL--QGTPTQOMAKSPKEMVQTPREYKAD 2988
QY      1463 ASTGSKGHVBSLIGSPGRTFPPVNPIDVADAPALERACYESLSKRP-----GTASSS 1518
Db      2989 LSAEGENDE-----SPSSPNNMSLHPLD-----QELSLSSQPHGMIPHNPT 3030
QY      1519 GGSII-----ARGAPVIVPEL-----GKPRGSLTYEDHGAAPFAGHLPRGSPVTRE 1564
Db      3031 HGKIYLHYAEPPTGPFVEPRDLFLTKTKSKPVQSGPSQIDK-----SKPEVFS 3079
QY      1565 PTPRLQEGSLSSKASQ-----DRKLTSPREIANKSPHSTVEHNHPH--PIS 1609
Db      3080 QSPSEBESVLAQAQGEESRAPRPMMSLOLDDQKLT-----LSSHPHMIF--HNHPNTEGK 3133
QY      1610 PYEHLKGVSG-----VDLY-----RSHTPLAFDPTGISIRG-----IPDA 1645
Db      3134 IYLHYAEPPTGPFVEPRDLFLTKTKSKPVQSGPQVDSKPEMFTQSPYEESVLAPOA 3193
QY      1646 AA--AYULPRHIANP-----TYPH--LYRP-----YLIRGVDTALLENROQII 1666
Db      3194 EGQESRAPRPMMSLOPLDQDLTLSSHPHGMITHHNPTPGKIYLHYAEPPTGPFVEPRDL 3253
QY      1687 NDYITSQCMHNHTATAMQADMLRGLSPRESSALANYPAGRGIIIDLSQVPHPLVLP 1746
Db      3254 FLRTTKSKPAQMTPTQMAKSPBEMVSLSPENKETY--PPAQGQGEESISPRHMSLOQLD 3311
QY      1747 TPGPAPATMDRLAYLPTAPQPSG--RHSSSPLSGG-------THLYKPTTSS 1793
Db      3312 QDLTPSSHPH--GWIPIHHPNTHGKIYLHYAEP--PTGPFVEPRDLFLTKTKSKPVQSP 3367

```

```

QY      1794 SERERDRERDRERERKSLTSTTYVHNPIWPRPTGEOSSGSSGGGSSSRPAS 1853
Db      3368 ROIDSPEKEVFTQSEYEVSL-----PAQEGEE----- 3398
QY      1854 HSHAQSPISPRTD--ALQORPS--VLHNTGMKGI-----TAVESKPTVRS 1900
Db      3399 -SRAPRMSLOPLDQDLTLSSHPHGMITHHNPTPGKIYLHYAEPPTGPFVEPRDLFLKT 3457
QY      1901 TS-----TSSP-----VRPA-----ATPP-----A 1916
Db      3458 TSKRPEHSGPRQIDSKHEMFTQSEYEVSLPAQEGEESRAPRPMMSLOPLDQDLTLIS 3517
QY      1917 THCRLG-----GLIDGYPTLMEVALLP-----KEAPRARERPRADTGH 1957
Db      3518 SH-PHGMIPHNPTPGKIYLHYAEPPTGPFVEPRDLFLTKTKSKPVQSGPRQIDSKPE 3576
QY      1958 AFLAKPRASGLEPASPSK--GSEPRPLVPVSGHATARTPAKNIAPRNA--SPD--- 2010
Db      3577 MFTQSPYEESVLAQAQGEESRAPRPMMSLOPLDQDLTLSSH--HGMIPHNPTPGKIY 3635
QY      2011 -----PPAPPASADP--HREKTSKPFSLQELSLSLGYHS-----SYSP 2050
Db      3636 LHYAEPPTGP--FVEPRDLFLTKTKSKP-----VHSGPRQIDSKPEKEVFTQSP 3682
QY      2051 EGVEPVSVSSPSLTHDKGLPKH--LEBLDK-----SHLEGEL--RPKQRPVULGGBAA 2101
Db      3683 EYESVLAQAQGEESRA--PRHMSLOPLDQDLTLSSHPHGMITHHNPTPGKIYL----- 3736
QY      2102 HPLRLPLRESQPSRPLQTPAGVGHQAVVTLAQHISEYITODYTHNHPQOULSAPRA 2161
Db      3737 HYAEPPTGPFVEPRDLFLTKTKSKPVAGSPRQIDSKPEVFTQSG--PEYEVSLPA 3791
QY      2162 PLYSPFGASCP-----VLDLRPPSLYLP 2189
Db      3792 QAEDQESRAPRPMMSLOPLDQDLTLSSHPHGMITHHNPTPGKIYLHYAEPPTGPFVEPRD 3851
QY      2190 -----HGAPARSP-----HSEGKRSPEPKTVS--LGGED-----GIEPV- 2225
Db      3852 LFLRTTKSKPVQMTPTQIDSKREBMSQSPENEBADIPRHGDQDELNPPIISLOPLD 3911
QY      2226 -----SPREG--MEPGRSRAVYPLLYRD--GEOTEPHMSKSPGNTSQPAFFSKL 2275
Db      3912 QELTSSQPHGMVPHNPTPGKIYLHYAEPPTGPFVEPRDLFLTKTKSKPVQSGPKELA 2988
QY      2276 TESN-----SAMYKSKQELINKKLTNNHNEPYN--ISQGTETFMNPAITGTGLMT 2326
Db      3958 TSKRPVQGTPTQIAKSPKEMVFO-----TPREKAVLSAPGEDODETP----- 4000
QY      2327 YRSQAVQENASTNNGLEAIIRKALMGKYDQ--WEESP-----LSANA-FNP----- 2370
Db      4001 -----SSPNNMSLOSLDOELTMSQPHGMILHNPTPGKIYLHYAEPPTGPFVEPRD 4051
QY      2371 --LNASASLPAMPITAAADGRSDHTLISPGGGAAYVSGRPSRKAAPAGLASGDRP 2427
Db      4052 LFLTKTKSKPVQGTIVQMD-----KSPKQGLL----- 4079
QY      2428 PVSASVHSEGCNRTPLTNKWE--DRPSAGSTPF-----PYNPLIMLQAGVMAAP 2480
Db      4080 ---PAQEAQDESRSPNNMSLOPLDQDLTLSSHPHGMITHHNPTPGKIYLH---YAEP 4132
QY      2481 -----PPGL-----PAGSGPLAGPHNMADEPRKLLSQVETLSDSE 2517
Db      4133 TGPVEPRDLFLTKTKSKPLAGITPTQOMAKSPKEMFS---QTEHNKE 4176

```

RESULT 9

S36152
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human
 C:Species: Homo sapiens (man)
 C:Date: 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change 15-Sep-2000
 C:Accession: S36152
 R:Itis, F.J.M.; Bouguenelret, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Jurka

Db 1935 -AVALAPQTVPEVKDTSKALETLTVSPAKSGDCLHSPKGVSVQVATPLLAFTSDKVP 1992

QY 2374 ---SASLPAMPITPAADGRSDHLLT---SPGGGKAKVSGRPSRKAKSPAPGLASGDRP 2427

Db 1993 EAVASAV-APKPADPA-----SLTLASP-----VAPLPKPKPLLESAPG--SVLES 1936

QY 2428 PSVSSVSHSEGCNRRTPLTNRVWEDRPSASSTFPNPLIMRQAGVMSPPPPGLA 2486

Db 1937 PSKLPVPAEED-ELPPLI-----PPEAVSGGEPPQPLLVN-----MPAPKAGTFA 1981

RESULT 8

A41819

proline-rich peptides 637K precursor, prostatic - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 13-May-1992 #sequence revision 13-May-1992 #text_change 05-Nov-1999

C:Accession: A42663; A41819; A31966; B20593; A20593

R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.

J: Biol. Chem. 267, 9884-9894, 1992

A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the leu exon.

A:Reference number: A42663; MUID:92250652; PMID:1577819

A:Accession: A42663

A:Molecule type: DNA

A:Residues: 1-3706,'1',3708-4077,'F',4079-4155,'S',4157-5761 <DE2>

A:Cross-references: GB:M86514

A:Experimental source: ventral prostate

A>Note: sequence inconsistent with the nucleotide translation

R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.

submitted to the Protein Sequence Database, April 1992

A:Reference number: A41819

A:Accession: A41819

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-5762 <DE1>

R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Wind

J: Biol. Chem. 263, 19159-19165, 1988

A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich F

A:Reference number: A31966; MUID:89066721; PMID:3198617

A:Accession: A31966

A:Molecule type: mRNA

A:Residues: 3372-3540 <HEM>

A:Cross-references: GB:M20721; GB:J04188; NID:G206397; PIDN:AAA41950.1; PID:G554494

R:Peeters, B.; Heyns, W.; Bossyns, D.; Rombauts, W.

J: Biol. Chem. 258, 14206-14211, 1983

A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primary s

A:Reference number: A94675; MUID:84061859; PMID:6685733

A:Accession: B20593

A:Molecule type: protein

A:Residues: 2020-2057 <PEB>

A>Note: this peptide, designated proline-rich polypeptide V, can be found at several loc

A:Accession: A20593

A:Molecule type: protein

A:Residues: 2822-2859 <PE2>

A>Note: this peptide, designated proline-rich polypeptide IV, can be found at several loc

C:Genetics:

A:introns: #status absent

A>Note: single copy gene with no introns

C:Superfamily: rat prostatic proline-rich peptides 637K precursor

C:Keywords: prostate; tandem repeat

F:1-56/Domain: signal sequence #status predicted <SIG>

F:27-5761/Product: prostatic proline-rich peptides 637K proprotein #status predicted <M

Query Match 3.9%; Score 512; DB 2; Length 5762;

Best Local Similarity 18.9%; Pred. No. 2.9e-10;

Matches 609; Conservative 342; Mismatches 1164; Indels 1112; Gaps 155;

QY 42 YOHHSRDVASHLSPSSIIOPRRRPSLSLSEFQNGRSGQELHLRPES--HSLYDELGKSE 99

Db 1311 YQVSGQDAQAEYTTSTV-----SRP---LDQELITLSEALREHNHPRVPROT 1355

100 MFIESKRPRLLEPDLPLRSP-----LLATGQAPGSEDL--TYDRSLITGKLE 146

Db	1356	ITNHPTRHP-LVHSEQOQHNP	FEVTVVQRIDELMTTPORTAGELRQTLQDSPTQIIE	141		
QY	147	PVS-----	PRSPHPTDPELEL-----	163		
Db	1415	PRTVVGPVPIVEEVVQTTSSODQAEVBPSPFVS	FOISDGLGLITTPBETHPITQKTVF	1474		
QY	164	PRRLSKELLON-----	MDVDRITMVGEQOISLKKKQOOLEBAKPREPE	211		
Db	1475	PRMVTDTLPRQVSVHKLKTEGIVQRIDELITTPORPTE	GEISQVQDSTTONKEPH	1533		
QY	212	KP-VSPPIESKHSVLVOIYDENRKKAEAHRILEGIP	QVELPLYNQSDPTROYHEN	269		
Db	1534	KEVAVAPRV-----	YQAVTVPIRPSQYQAEYC-KSLQRIDELITYISEPT-KEAVHST	1583		
QY	270	IKINQAMREKLLIYFKRNHARKQWKQKFCORYD-----	QIMEALEKVEV	314		
Db	1584	ISKSLAINQYVHIOHPNPA-----	EKVQRIPLDELITSSLSLOPTAGELLISMQEIVT	1638		
QY	315	RIENPPRRRAESKRYREYKQFPEIKQRILOERMGRVQORSGLSMSARSHETSE	374			
Db	1639	QISEPPKQV--TVPEPEQEVAVP-----	APVQOQAKPL--SSIYSLSLDOGLTSS	1688		
QY	375	IIDGISEONLEKQMRQOLAVIPM-----	LYDADQORIFIMNGGIMADPMKYKQRYVN	430		
Db	1689	ELIIEAHOLITPD--	ETMVLPRDQIGIT-PRDHDKHMLDTEVTNORPHLEHTVQOR	1744		
QY	431	MMSEOEKETPEREKPMO-----	HPKNFGILASFLEK	461		
Db	1745	TIEBERSQSLOKTTQTUTEPRKQVVP	PLAQSEEVITIPWILKETAFPRPHSMALQSLDEK	1804		
QY	462	-----	TVAEVUY-----	YLTKNEN	478	
Db	1805	LTISHSPGWTQOAHNLKESKHTTGKLDIYAEBNNEI	BLKHGFLFKLTTEATTESEN	186		
QY	479	YKSLVRRYRRRGQSQQOQQOQQOQO-----	QQOQQOQPMRPSQOEBKEKEKEAEKEEBEK	536		
Db	1865	TNQMTKSLIKQYTALFTQNKXSMPLALVBSQDESQRP	PMMSIQRIQDQ-ELTSSQCHGVP	1922		
QY	537	EVENDKEDLLEKTDTSGEDNDEKEAVASKRTANSQGR-----	KGRITRSMAN	588		
Db	1924	HIPMTPEKITIYHVAEPRTGPFVEPPELFLKTTKSR	PVQGTATQMAAPKEMVASPAENK	1983		
QY	589	-----	EANSEAITQOQAEIAMSMLANSSR--	WTEEBEMETAKGLEHGNMMAI	637	
Db	1984	EAVLSGPEDODESPSPNMSIQSLDQELTSSOPHGI	PHRPNT-----	HGKITIYH	2036	
QY	638	ARMWGSKTVSOCKNFYFVKKRONDELIOQHKLM	KEKRNARRKKKAPAAASEAAP	697		
Db	2037	AEPTGPFVBRPDLFLKTTKSKPVEWTLTR-----	TDKSRKEMVSOQKTYEAVLP	2088		
QY	698	PVDEDEEMEASGVSGNEEBMEAEALHASGNEVPRG	CSGPATVNNSDTESIPSTE	757		
Db	2089	-----	VHGEQOEBER-----	SPNMSIQSLQOE	2111	
QY	758	AAKDTGONG--	KPRPATIQA-----	DGPPRGPRTP--	KRTSRPIEPT-----	797
Db	2112	LTLSSQPHGWNPHRPNTIGKITIYHVAEPRTGPFVBRPDL	FLKTTKSKPVQGTATRMVKS	2177		
QY	798	-----	PA-SEATG-APRRP-----	PAP-SPS-	816	
Db	2172	PEEWVSLDPENKEAVFPAQGEQKGSFSPNMSIQSLDHEL	FMSQORPHGIMPHRPKITPDK	2233		
QY	817	-----	APPVVPEKEKEE-----	ETMAAPRVEGEQKPPRAAEIADVTKAEVYKSECT	867	
Db	2232	IYLNHAEPRTPGFVBRPDLFLKTTKSKPVQGTTEMAKSP	REWVSQTPREYKEAVLSQSG	2299		
QY	868	EEAEERGA-----	KGXDAEALEAT-----	AEQALKAEEKGSGR-----	902	
Db	2292	EDODESPRPMTSLKSLDQEVAMSSQPHGVNHRPRTGKITIYH	SIERRPPGPFVVPRTDI	2351		
QY	903	--ATTAKSSGA-----	PODSOSATCSADDEVVBAAGDQKNRLSPRP	942		

Query Match 3.9%; Score 520; DB 2; Length 2187;
 Best Local Similarity 19.7%; Pred. No. 5,4e-11;
 Matches 448; Conservative 219; Mismatches 782; Indels 830; Gaps 93;

QY 738 GPATNNSSDTEISIPHTAA-----KDTGQNGPKPATLGADGPP--PGPPT 784
 DB 3 GEATVATTEBELPOQAFETAVLPMSSALKVAAVGQGPPTPSSILGQGSIVTAHOPS 62
 QY 785 P-PRRTSAPRT-----PTPASEA-----TGAPTP-----PP----- 810
 DB 63 PLFSSVSTPFEVPPAQTITTAETLPSGTAPPTPTPLPHLICPPISPAALALASPMIGLA 122
 QY 811 -----PPSPSAPPVVPEKEKEETAAPVE----- 838
 DB 123 QKARSSAPLSLVALPHSVQSSVCPPHPLTSPPSAAGALGLTMSIPLEPKTSTS 182
 QY 839 -----EGEEOKPP-----AAELAVDTGKAEPVKE 865
 DB 183 QVPSQGTINLKGTACPPDVVAFPSHLENPLASVQPGIMSCPTLSMTSPVKGVPISSA 242
 QY 866 CTBEABEGPAGKDAEAAETAGALKAE-----KKEGGSGRAT 904
 DB 243 LTQSLSLNLKGPVPPRANTAPASIPLAPSTSLCCHPLHSSVDSPIQPGQSGLAIV 302
 QY 905 T-----AKSGAPQDSDSATCSADEVDEAGCDK- 934
 DB 303 SNPTVGHSGIAACGPBRCVVPALPSRLAVDSGAAPSDDKSSAVTNELCSPGSSNV 362
 QY 935 -NRLLSPRLTTPGDPRAANASFOKLDLKQKRAALPIPIQVTKVHEPP----- 985
 DB 363 AGTSLSPKASLV-----PKGSNVALQPL-----VTQVPASQCTGKEIFVSCIGATH 409
 QY 986 -----REDAPTPAPAPPPQNLOPESDAPQPGSPRKSRSAPADKFAFAEAKOX 1041
 DB 410 HALNPSALISVAPATHVPPPT-----SGGLVSKDPSAPVTSLVPPAAKQ 455
 QY 1042 LPGDPCWTSGLPF-PVPPREVITKASPHAPPSAFSVAPEGHPLGLH---HDTARFVL 1096
 DB 456 FPAPASATLVGPVPEPLPATEGLK--NLPIALVNGAVPSPAQAGLPTRKDTTLQOLA 512
 QY 1097 P-----RPPTISNPPLISSAKHPSVLEROIQAISQMSVOL 1133
 DB 513 PIALKESPSQSASSEVLSEDTVTYKTTGGPAPVVRPA-----IAGVATTTISARA 563
 QY 1134 HVPYSE-HAKAPVCGVTWGLPL-----PMDPKLIAPSGVQKQOLSPRGAGPPE 1182
 DB 564 DSPRAVTAADCGSVSNVTISQPLKSVTPPAMAFTAKTAPST--TSPVLPLASEGCPV 620
 QY 1183 SLGVPTAQASVLRGTALGVPGGSITKQIP-----STRVPSDAITTYRCSI 1229
 DB 621 ASSMALSPONASVSFTALALSP--EIPKSVPPDPPLAEIFSNARKVDVAVSHMSSGSS 678
 QY 1230 THGTP-ADVLYKGTITRIIGEDSPRLDRGRDS-----LPKHVIT 1269
 DB 679 RQGHDPDASVYAKGVVCL-----ADSLDTSVASASGALSASGASPLYLEVSLPEAGLA 734
 QY 1270 YEGKGHLVYEGGMSVTQCKEDG--RSSGPPHETA----- 1306
 DB 735 VQGRKGL-----NKLSTPPSSKGAAPVSTGAPPKAPIVPTTSSISSQVPAELLPS 790
 QY 1307 PKRTYDMMEGRVGAIRASISIEGLM-----GRAIPPERHSPHMLKEOHNT-- 1351
 DB 791 FQKPEVYASRLISAVQSPKVDPIMSDVTPTSPKKTSAVAPKOTSATLSKSVAVTSL 850
 QY 1352 -----RGSITQGIPIR----- 1361
 DB 851 SPKAPVAPASNEATIVPTPIPTSLKNALAAATPKETLATSIPKVTSPQKTPKSVSLKG 910
 QY 1362 -----SYVEAOED-----YLREAKLTKRE-----GTPPP--PP 1388
 DB 911 APAMTSKATETLAASKDVSPSQPFKEVPLQHVPTTSPKPSVSDTLTSGALTSPPPKCP 970
 QY 1389 PSRDLTEAVYKTOALGPLLK-----PAHEGLVATVKEAGRSIHETPR-----EELRHTEPEL 1439

DB 971 ATLAEPTTPPKSPKAPASKKTTPATPSPEGVTAVLEIPCKSKAPKTAAPKESATSSS 1030
 QY 1440 PLAPPLKEGSIITQ-----GTLKYTTGASTTSSKKHNDVSLIGSGRT----- 1483
 DB 1031 KRAPKTAASKEIPSKGVTAVLEISLPLKETSKSATPDEKS-----ASSPKRSKPTAGP 1084
 QY 1484 --FPP-----VHPDVMADALERACVEESI-----KSRPGTA-----SSGGGSIAR-- 1524
 DB 1085 KETPPGVTAVPPELSLPPKKTPONATPNESILASSQKRSKTSVPKKTPPGCVTAMLE 1144
 QY 1525 --GAPVIVELQKRPQSPDYEDHGAFFAGHLPRGSPVTMBEPTRLOEGSLSSSKASQD 1582
 DB 1145 IPSAQKAPKTAAPVQPIR-TPED-----AVTLIAGSLSPKK-----ASKTAAP 1187
 QY 1583 RKLSTPR-----ELAKSPHSI-----VPHNHPPISPYEHLRGVSGV 1621
 DB 1188 KEAPATPSGVIAVSGEISPSPKKTSKTAAPKENSATLPPKRSKPTAAPKETPATSSBGV 1247
 QY 1622 DLVSHIPLADPISIPRGIFLDAAAAYLLPRHLAPNTYPHLYPPLYIRGVPDTAALEN 1681
 DB 1248 TAVSELSPS-PPTPASKGVV-ITTPKGAVALAESPASP-----KVPKTAABEE 1297
 QY 1682 RQTIINDYITSOQHNTATAMAQRADMLRGLSPRESSIALNVAAGPRGIIIDLSQVPHLP 1741
 DB 1298 TST-----TSPQKIPKIVAG--PREAS----- 1317
 QY 1742 VLVPPTGCTPATMDRLAYLPT-----AQPSPSRHSSGPLEBPGFTHLTKPPTT 1791
 DB 1318 -ATPPSKTPTTAVPKETASBEGVAVPLEIPSPRAPAPTAAP-----KETPA 1366
 QY 1792 SSSERERDRDRDREREREKSLTSTTVVHAHILMPRGTEQSSGSSGSGGSSSRP 1851
 DB 1367 PSPEG-----ATTAPVQIIPSPKGSKA-----GSKETP 1386
 QY 1852 ASHS-----HAHQSPISPRQDALQORPSVLAHNTGMKGIITVAPEKPYLRSTSTSP 1906
 DB 1397 TTPSBEQVTAAPLEIPISSKTSKMAKPEKTLVTPSSKSLQVGPKEKTSLEGATVAVLE 1456
 QY 1907 VRPAATPPAT-----HCPGLGTLQGVPTLMEPVLLPREARVARVPERPRADTGHAPLAK 1962
 DB 1457 IPESHKAPKTVDPKQVPLTPSPKDAPTTLAESPSSPKAKTAAPESEKVTY-----V 1510
 QY 1963 PPASGLEPASSPKGSEPPRLVPVPSGHATITAPKAKNLAPHNASPPAPPA--SASD 2020
 DB 1511 PP-----EKATPQKASGTTASKVPV-----PAETGVANASSKETPTTAPVPVKN 1556
 QY 2021 PHREKTSKPSIOLELRLSLGYHSSVSPBGEVPSVSPSLTHDKGLPKHLELDKS 2080
 DB 1557 PSHKKTSTKTELKE-----APATLPP-SPTKSPKIPSSKAPR----- 1594
 QY 2081 HLBELAPKQGPVYKLGGEAHLPHLRPLPS--QPSSSPLLOTAPGVKHGRVVTLAQHI 2139
 DB 1595 ---TSAPK-----EPASPSIKPVTTISLAQTAPSLQKAPS----- 1627
 QY 2140 SEVITQDYTRHHPOLASPLPAPLYS--FPGA-----SCP 2172
 DB 1628 ---TTIKKNLAAAPALVPSKSPAPAPASASLSAPATAQCAPKEATTIISCK 1679
 QY 2173 VLDIRRPPSDLYLPPDHPAGARSPHS-----EGGKRS-----PEPNK 2211
 DB 1680 KAAATEPIETNSTAPLEGAPKETSETSVSKVLMSSPPKXASSKRASTLPATLPLKE 1739
 QY 2212 TSVL-----GGGDGIEPVSP--PBGMTEGHRSARVAPLYLYDGGQTEPSRMGSSSPG 2263
 DB 1740 ASVLSPTATSGKOSHISPVSDACSTGTTPQASEKL-----PSKKG----- 1781
 QY 2264 NTSQPPAFPSKLT--EANSNAVSKKOEINKKNTAHNRNEBEVNIISQPTEIFMPPAITG 2321
 DB 1782 ---PTAFTEMLAPAPESALATIAPIQKSPGANSNSASBPCKCDPSKDOTKLPS--- 1834
 QY 2322 TGLMTYRQAVOEHASTNMGIEALITRKALMGKYDQWEESSPLSANAFNPUNA----- 2373

```

Db      653 QOHQGSAPPTVPSPPOQVTL-GAVPAPQAP-----PPPKALY----- 692
Qy      1043 PGDPPTCSGLPFVVPREVIVKASPHAPDPSAFYAPRGHPLPLGLHDTARPVLPRTPT 1102
Db      693 PG-----ALGRPPMPMNPF-----DPRMMIIPVVD-----RLQGRPLE 730
Qy      1103 SNRPPLISAHAPSVLERQIGATISQMSVOLHVPYSEHAKAPVPTMGLPLPMDPKKA 1162
Db      731 FYRPGV-----HPSGLVPRERSDLSGLSE---PFDHAPAMML-ERGIT-PVD-KLA 778
Qy      1163 ----PFGVKOEQ----LSPRQAGRPESLGVPTAQEASVLKGTALGVSFGSITKGIPST 1215
Db      779 WVGQVFTATAPRAPPLPLSLRQADEDDKGM-----RSETPVP- 818
Qy      1216 RVPSDAITTRGSIITHTGTPADVLYKTTTRI-IGEDSPRLDREREDSLPGHVIYSGKK 1274
Db      819 -PPVYLASYGFPENAGAP- -DISRPLEEPGRPL- -PWPG----- 857
Qy      1275 GHVLYSGMSVTCSEKEDGSSSGPRHETAPRKTYDMEGRGVRAISSASIEGLMRA 1334
Db      858 ----SDEVAKITPPPKPEPKETQLTGPEAGKULPASR---SGAG 898
Qy      1335 IPRERHSPHLKEQHIRGSIITGIPRSYVAQEDYLREAKLIKREGTTP-----P 1386
Db      899 PPPRRASRTETRKPRRSGSRGIP-----EPGAPRRRAGPIKKP 941
Qy      1387 PPPRDLTEAVKTOALGPLKLKPAHEGLVATVKEAGRSIHEIPREBELHTEPLAPPL 1446
Db      942 PPPRK-----VEELPPKPLEOGDETEKPPRP- 967
Qy      1447 KESITGOTPLKYDTGASTGSKKHVRSLLSGPRTPP- -VHPLDVMAADALER- 1501
Db      968 ----DPLKITK- -LGGPKET- -PENGUISPAPRLRKYSEYEVG 1005
Qy      1502 -ACYESL- -KSRPTASS- -GGSITAG- -APVIYE 1532
Db      1006 PTCRGRGRGEYFARGRGFRGTGGRGCGQANSVTEFEMMGVEVGQDDQTLLPE 1065
Qy      1533 LGKP- -RQPLTYEDHGAPRAGHLPKRGSPV- -TMRPPLIOEGSL 1575
Db      1066 AAMPARGARVQSMKSP- -SGAGSGAQOKAPRPMRVIMLQTRLLPFRREHSPRS 1120
Qy      1576 SSKASQBRKLTSTRELAKSPHSYVPEHHPRPISYEHLLRGVSDLYRSHPL-APDP 1634
Db      1121 S- -RSPPTNSPTLHR- -APARTCPGV- -BSLEBGALSP 1155
Qy      1635 TSIPRGIPLDAAAAYLP-RHLPN- -PTYPHLYP- -YLIRGYPTALLENROT 1684
Db      1156 GPRRREARPOVCPGMSPPAKSLAPKXPTGP- -LPSPKEPLKXLIPLPLPVARCGSNG 1213
Qy      1685 IINDYITSQMHNTATMAQPADMLRGLSPRESSLAL- -NYAAGPRGITDLSOVHLP 1741
Db      1214 GSNVGMEDGERPRRRRHGRAQOQDK- -PRFRRLKQERENAAAGSEG----- 1258
Qy      1742 VLVPPTPTKTA- -MDRLAYLPAPQPFSSRHSSSLSGCPHLLKPTTSS----- 1793
Db      1259 - -KPSLTLPAASQPEBALTTVVAPAPRAAKSPDLSNONSQANEMETSESSDF 1315
Qy      1794 -SERERDRRE- -RDREERKSSILTSTTT 1820
Db      1316 TSERRGDEARPPVILTPKAVGTGCGGAGVPGISMSRDDLQGRADLKRSSSGRP 1375
Qy      1821 VEHAPIWRPTEGSSGSSGSGGCG- -SSSPASHSHAHQ 1859
Db      1376 GMEQRNRRPGCGKXGSSGSSGSGGCGGPGRTGGRGDKRSPKRNRSRPPERRPGL 1435
Qy      1860 HSPISPRQDLOORPSVLHNTGMKGITTAVEPSKPTVLASSTSSPVPRATPPPATNC 1919
Db      1436 PLPPPPSSSAVFRIDVYH- -SNPAGIQALQSLRQGSVTAPGCH- 1481
Qy      1920 PLGGLDGVVPTLMEPVLLPKAPAPVAPRPRADTGHAFKAPAR- -SGLEPA 1972
Db      1482 - -PR-HKGPQAPQGS- -PRPRTREPORVNSGL- 1512

```

```

Qy      1973 SSPSKGSEPRILVPSG- -HATARTPAK- -NLAPHAS----- 2008
Db      1513 SSDPHEEPQPMVRGCGTTRDQAGVSPPPKRRERPRKRELLQESLPPSHSSGLGS 1572
Qy      2009 -PPAPAPASAD- -PH-----REKTOSKPSIQ-----ELE 2037
Db      1573 KPQEPQASRDTGTCTALPHIMNLHTATSKRSYRPTSMEMWEDLSPFEDVAGTEMS 1632
Qy      2038 LRLSLGHGSSYSPGCVPPVSPSSILTHDKL- -PKH-----LELDKSHLEG 2084
Db      1633 QSDSGVLDLSDSQVSSGSCQSRSSP- -DGLKGAEGPKRPGSSPLNAPCEBPG 1688
Qy      2085 ELAPKQCPVKGGEAHLPHLPLPESGSSSPLQTAGVGHORVTLAOHISEVIT 2144
Db      1689 SEPRRRPPAPHDGRKRELPBOLP- -DPIGTER- -SOR-----T 1726
Qy      2145 QDYTRHHPQOLSAPLPAPLYSPGASCPLDLR- -PPS-DL 2183
Db      1727 DRGTEPPIRPS-HRPGPVQF-GTSDKSDLLVGDLSLKAKELTASVTBAIPVSRDW 1784
Qy      2184 YLPPDHGAPARQSPHS- -EGKRSPEEN- -KTVLGGEGDGT- 2222
Db      1785 ELAP- -SAAASAEPQSKNLDGCHVPEPSSGQRLYPEVFGSAGPSSQISGSGSHLS 1841
Qy      2223 - -EPVSPKMTBPGHSRAV-YPLLYRQGEQTEBPMGSKS 2261
Db      1842 ITSQKMLRPGTSLPHYSQPLPLPPGAPPSALLSGVALKQGLDPTMQATELQKL 1901
Qy      2262 PGNTQPPAPF- -SKLTESNAMYKSKQOEINKKLTINRNEPXYNISQPGTEI 2313
Db      1902 AGGVLYPPSPFLVSPACPSPLPDTSLQVRQ-----DLRSP-SDF 1941
Qy      2314 FNNPATIG- -TGLMTYSQAVQEHASTNMGLEAITRKALMGKIDQWEEPPLSANANPL 2371
Db      1942 YSTPLQGGSGGFLPSGAPAQO- -MLPMDSQLPVVNFGLSPAP- -PAPPL 1992
Qy      2372 NASASLPAAMPTAADRSHTLTPSGGGKAKXVSGRPSRKKASPPAGLASGDRPSPVS 2431
Db      1993 SLPLVGPALPPSLA- -VRPPAPPAITRVLPSP- -RPPPAS 2029
Qy      2432 SVHSE- -GDCNRTPLTNVWEDRPSASGTPPYNPLI- -M 2469
Db      2030 LGRALHLPVELKPRQDYQKLSNNGRSGSRTPTGFSFGLNSRLKATSTYSGVVRTQ 2089
Qy      2470 RLQAGVNASP- -PPGLPAGSGPLAGPHHAWDEP 2502
Db      2090 RVDLYQOASPPDALRWIPKPEWERTGPPRREGPSRR-AEEP 2128

```

RESULT 7

```

T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Yocov, W.V.; St-Arnaud, R.
Gene Dev. 10, 1763-1772, 1996
A>Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Reissues: 1-2187 <YOT>
A:Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Intons: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

```


Db DEAKSKESRESVAEQP-----LVSEVSRPASVAESVVDDEAKSKESBPLMSKASR 3149

QY 1776 PLSPGPPHLTKPTTSSSERERDRDRDREREKSLTSTTVEHAPIWRDTEQSS 1835

Db PASVAG-----SVDEAKSKESBR-RESVAEKSPLPSKASRPASVAESVKDEAD 3150

QY 1836 GSSGSGGGGSSSRPASHAHQHSPISPTDOLQ--QRPVLAHTGMKGITTAVERP 1893

Db KSKESRRESGAEKSPPLASKASRPASVAESIKDEAKSKESRRESVAEKSPLPSKAS 3259

QY 1894 KPTVL-----RSTSSVPRPAATFPATHCPLGTLIDGVYPTLMEVLLKKEAPR 1944

Db RPTVAASVKDEAKSKESSRDVAEKS-----LASKASR 3297

QY 1945 VARDEPRADTGAFLAKPARSGLEPASD---SKGSEPRPLVPVSGHATTIARTAKN 2001

Db PASVAESVQD--EAKSKESRRESVAEKSPLAYKASRPASVAESIODEAKSKESBRR 3355

QY 2002 LAPHNASDPAPPASASDPHREKTQSKPSTIQEILRLSLGHSSYSVPEGEVPSVSS 2061

Db 3356 -----ESVAEKSPLASKASRPSTVAESVKDEAKSKESSRDVAEKSPLAS 3403

QY 2062 PSLTHDKGLPRHL--EELDKSLHEG--ELRKQRPVVLGGEAHLPHL-----R 2107

Db 3404 KEASRPASVAESVQDEAKSKESRRESVAEKSPLASKASRPASVAESVKDEAKSKEE 3463

QY 2108 PLPESOPSSSPLLQTAPGVKGHQRVVTLAHISEVY--TODYTHHPQOLSAPLPAPLYS 2165

Db 3464 SRRESVAEKSPLAS-----KEASRPASVAESVKDEAKSKESRRESVAEKSPLPSK--- 3515

QY 2166 PPGASCPVLDIRRPPSDLYLPPPDHGAAPAGSPHSEGGKSRPEPNKTSVLTGGEDGIEPV 2225

Db 3516 -----EASRPST-----VAESVKDEAKSKESRRESVAEKSILASKEA 3554

QY 2226 SPPEGMPFGHSKRAVYLVLRDGGQTEPSMGSGSPENTQOPAPFFKLTENSAMVKS 2285

Db 3555 SRPASVAE-----SVKDEAKSKESRRESVAEKS-----LASKASRPASVAES 3600

QY 2286 KKOELINKLANTHNRN-----PEYNISOPCTEIFNMPALTGTGLMYRQAOVQEHAST 2338

Db 3601 VKDEAKSKESVSRRESVAEKSPLPSKASRP-----TSVAESVKDEADK 3644

QY 2339 NMGLEAIIIRKALMGKYDQWEEBSPPLSANAFNPLNASASLPAAMPITTAADGRSDHTLSPG 2398

Db 3645 SK-----EESRRRESGAEKSP---ASMBASRPSTVAESVYDET----- 3679

QY 2399 GGGKAKVSGRPSSRAKSPAPAGLASGDRPPSVS--SVHSEGDGCR-----RTPLTN 2447

Db 3680 --EKSKESRRESVTEKSPPLPS-KEASRPSTVAESVKDEAKSKESRRESVAEKSPLAS 3736

QY 2448 RWMEDRPSSAGST 2460

Db 3737 KE-SSRPASVAES 3748

RESULT 5

TS1023

hypoetical protein B7F21.40 [imported] - Neurospora crassa

C/Species: Neurospora crassa

C/Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C/Accession: TS1023

R/Schulte, U.; Aign, V.; Hohlsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, July 2000

A/Reference number: Z25286

A/Accession: TS1023

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-2649 <SCH>

A/Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40

A/Experimental source: BAC clone B7F21, strain OK74A

C/Genetics:

A/Gene: NCSP:B7F21.40

A/Map position: 6

Query Match	4.2%	Score 552;	DB 2;	Length 2649;
Best Local Similarity	19.3%;	Pred. No. 4.9e-12;		
Matches	406;	Conservative 268;	Mismatches 771;	Indels 658; Gaps 92;
QY 60	QPORRRLSEFGPG--NERSQGLHLPRESHSLPELGKSEM--EFIESRRPRLLELPD	115		
DB 848	KPVKPKAQKLDLEPPRVSEBPVAPRPEEPKVESEPKKEKAPELLPELPEPEQIPR	907		
QY 116	PLRPPSP-----LLATGQPGSEDLTKRSLTQKLPSPVSPSPHTDPE	159		
DB 908	PQRPQRPQRPQQAQAQAPRQLLTALPEKSVTEPERERAEAEVQEAETQEQAP-TERQ	966		
QY 160	LELVP-----PRLSEEL-----IQNMDRVREITMVEQISKL	193		
DB 967	LLAIPVTEENETSFLRPREPQPKHEMDTGSGSLPLPTEVHKHVQDDVMDV-----	1020		
QY 194	KKKQOQLEEE--AAKPEPEKVPSPPIESKHSVQIITDENKRAEAAH-----	242		
DB 1021	-----VDDEGRDLQRP--PCKPVPILRDVSGDSGVSQFPPEFQOPTSDGASSRMEDS	1072		
QY 243	-----RIIEGLPQVELEPLYNQPSDTRQYHNIKINQAKRK--LILY	283		
DB 1073	EEETEDDASVSGSVENHVELSTTPTEDELPAFVN-----VKPMHQSCKRAKLSQSGPGRPAF	11228		
QY 284	F-----KRNHARKQMKQKFCQRYDQLEALKEKVEIEENPRRAESKYRE	331		
DB 1129	FMGNIRDAEETAKEQHEA-----KQVRYTNDAYLRF-----TLSDDP--TAVKSR---	1173		
QY 332	YEEKQPEPIRKQRELOEFMOS--RVGORGSGGLSMASARSHVESEI--DGLSE-QENLE	386		
DB 1174	---NQFSQSDKDXGSGKGSHSGSDNHGEGGRRRTTSRETELDEYVALQESIREAQEKKE	1230		
QY 387	KQMR-----QLAVIPMLY--DADQRIKFINMGMLAMDPMKVYKDRQVMMW----	432		
DB 1231	KEERQREKTYTDKEAVIPEMITWDEEDHOLEFDTSL--PLV-----KLIVETQAVP	1283		
QY 433	-----SBOEKTFRFEKFWQHPKNGFLASFLERKTVAECVLYYLTKKNENYKSLVRSY	487		
DB 1284	YHVNATAEAEKFEKAYILEYFKQMGKLAHLEPNDHSHVIOYUAKKREMLKRELKQRP	1343		
QY 488	RRRGKSQOQOQ-----OQOQOQOQOQOQPMFRSSQOEKDEKEKEAEKEBEKEPV	538		
DB 1344	RRRKKGKQKQKYNALVSELGNPENETTEDNGENGENSNGRQOPRAAAPMGHEATENA	1403		
QY 539	END-----KEDLKEKTDPTDSEDNDEKEMAVASKRKTAANSGRKKGI-----	582		
DB 1404	DSDGATPSATGRRRAAGTTTTPPKDSCAE--KEBGVKGGRRARQRPADKEPKVPKPAQ	1460		
QY 583	-----TRSMANEANSSEALITPOOSAE-----	603		
DB 1461	AIAPTPPAAASKGTNARSRNSTKVQNEBMPKTPALGLARVQNPMDVYPGSQPPLA	1520		
QY 604	-----IASMEINE-----	611		
DB 1521	PAQOTPLASPERAPPTLASTTISEWMAPPSLRPEBPAPASLPTFELIGSSGPERITRPQ	1580		
QY 612	--SSWTSEEMETAKKGLLEHGKMSALAAVNGSQTVSQCKNFFYFNKK--RONLDELQ	667		
DB 1581	QASSWSYSESNDFPSILRSFGTDSALANMGTTQVWVNNYYLRQKSGKEQWEQIAT	1640		
QY 668	QHKLMEKERNARAKKKKAPAAASEEAFPVVEDEEWAAGVSGNEBEMTEAEALHAS	727		
DB 1641	EADLK--KQGERRRPPPTPSAGR-----KRDVYSSSGHRLAAAEAE-----	168		
QY 728	GNEVGRGCGGATVNNNS-----SDTESIPSHTEAKD--TGQNGPKPATLGAD--GPPG	781		
DB 1685	-PQAKSEAAAPANOPFGRFQVTPRQASPVVHLTVQRPVTMPAPALPLPAPVAAASAPAS	1743		
QY 782	PPTPRRTSRAPLEPTPASEATGATPPRAPPSBPAPPVVPKBEKEETMAAPVVEGSE	841		
DB 1744	PQTPBR-----PGAPVQSTMGSVPHPL-----RQPAATVFTTERBEELI--PV-----	1784		


```

Qy 2085 ELRKQKQCPVYKLGEBAAHLPRL-----PLPESQSSSPLLQT--ABGVKQHNQVPLAQ 2137
Db 2021 ESTROYEP-----LHHYKQOQBSFSFOQPLPSPSSQSKMGQVPRTHRLITLAD 2071
Qy 2138 HISEVITQDYTRNH-POOLSAPL-----PAPLYSPFGASCPLDLRRRPSDLYLPPPP-- 2189
Db 2072 HICQITQDFARNOQVPSQASJSTGTQTSPLSLSTP-----VTKTSSKYSPESSQQT 2123
Qy 2190 --HGAPA-RGSPH-----SEGGKSPREBNKTSVLCGGEDGLEPVSPPRGMTIEPG-HKRSR 2240
Db 2124 VLHPRPGBRVSPENLVDKSRGSRGKSEKSEKH1--PSEPEPLSPGQ--PAVHEKQD 2177
Qy 2241 VYPLPYRQGEQTESRPMGSKSPGNTSCPPAEFSKLTESNAMYKSKQEIKNKKTTHNRN 2300
Db 2178 SMLLSQGVDPAPQRSDRSRPSGASIVLSPFTYL-BESTSMVYKSKQEIIRKLNSSGCG 2236
Qy 2301 EPEYINISQGTETINMPAITGTGLMTYRSQAVQEHASTNMGLEAIIRKALMGKYDQWE- 2359
Db 2237 DSDMAAQPGTEIFNLPVATTSGAVSSHSHPADAP--NLGLEDIIRKALMGSPDDKVED 2295
Qy 2360 -----SPLSANAFNPLUNASALPAPAMITAADGSDHLLTNSPGG--GKAVYSGRPSGRK 2413
Db 2296 HGVVMSHPV---GIMPGSASTSV-----VTSSEARRDDGEPSPHAGVCKPRLINKSNRK 2347
Qy 2414 AKSPAPAGLIA--SGDRPVSVSVHSEGCNRRPTPLTNRWEDRPSGASPPPNPLIMRL 2471
Db 2348 SKSPIPGQSYLGTERRPSSVSVHSGDYHRCQP--GMAMERPSSTGSTOPPNPLITRM 2405
Qy 2472 QAGVWASPPPGCLPAGSGPL--AGPH--HAMDEPRKLLCSQVETLSDSE 2517
Db 2406 ---LSTTPPQIOLCAPSAITQAAHQAONRIMERRPALVLSAQVETLSDSD 2452

```

```

Db      227  SQQWQVNL-----RPDVAGLSPREQRLGLPYPA-TRGIIDLTMNPP-TILVPHPGTS 278
Qy      1752  ATAMDRLAYLPTAQPPSSR-HSSSPLSFGCPHTLTPPTSSSERRDRDRDRDRR 1810
Db      279  TPRMDRTTYIPGTOITPPPPNSASNSPCHPTL-----AAASEREREKERERER 335
Qy      1811  EKSILTSTTVYEHAPIRMPGTGSGSGSGSGGGGSSSPASHSHHONSPLSPRQDA 1870
Db      336  ---LMASSDL-----YLRPGEQ-----PGRGSHGYVRSPSP-SVRQOET 373
Qy      1871  -LQGRPSVLHNTGKGIITAVEPEKPTVLNSTSSPV---RPAATPPATHCPLGCTL 1925
Db      374  MLOORPSFQGTNGTSVITPLDPAQLRIMPLPAGGPSISQGLPASYNTAADA-LALV 432
Qy      1926  DGVPFTLMEPLRLK-----BAPRARPERPRAD-----TGH 1957
Db      433  DAAASAPOMDVSKTKESKHEPARLEENLRBSAAVSSQOOLEKLTLEVERKSQVCLYTS 492
Qy      1958  AFLAKPPARSGLEPASS---PSKSGEPRPLVPVSGHATTARTPAKN-LAPHHASDP 2012
Db      493  AFPSGKP-----QHSSSVVSEAKDKGP---PPKSRYEELRTGKTTITPAANFIDVIT 545
Qy      2013  APPASADPNHEKTOQKPSIQELBLRSLGHWSSYRPEGVPEPSPVSSPLTHDKGLPX 2072
Db      546  ROIASDOKARERGQSSDSS-----SLSLSHRYETPSDALEVSPASSPPPOEKLQTY 599
Qy      2073  HLEELDSKSHLEGLERPKQPGFVKLGEAAHLPHLRPLPESQPSSSPLLOTAPG----- 2125
Db      600  QPEVVKANQANDPTRYOGEPP-----LHHYRP---QGESPSFOQLPPSSQAGMG 647
Qy      2126  -VKGHQVVTLAGHISEVITODTYRHH-----POULAPRLAPLYSPGA---SCPVLDR 2177

```

```

RESULT 3
T46489
hypothetical protein DKFp434M075.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46489
R;Dnestrheft. A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23035
A;Accession: T46489
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1047 <AAA>
A;Cross-references: EMBL:AL337641
A;Experimental source: adult testis; clone DKFp434M075
C;Genetics:
A;Note: DKFp434M075.1

```

[illegible]

RESULT 4
 T13564
 microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
 N/Alternate names: hypothetical protein EG:49E4.1
 C/Species: *Drosophila melanogaster*
 C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C/Accession: T13564
 R/Spinoe, L., Papagiannakis, G., Siden-Kiamos, I., Louis, C.
 Submitted to the EMBL Data library, April 1999
 A/Description: Sequencing the distal X chromosome of *Drosophila melanogaster*

Db	1	QSRYPHSNYQYTPRPAARHQOEPFVPRYBSHLEVSQAQSLUQOQOQOQOQRRRPSLSEPH	76
Qy	74	PNENRSQELHLNRPESHSYLPELGKSEMEPTBSKXPLLELLPPELLR-----PSPILATGO	128
Db	77	PASDRPOE--RRSGYQOFHGPSPVPHDSLESKRPLTEQVSDHFORIGAAVLPLVHT-L	133
Qy	129	PAG---SEULTDORSJLTKLE--PVSPSPPHNDPELTVPRPLSKELLONMDRVHRET	184
Db	134	PEGLSSANAKDPACGVHNEABSSLSGQPCDDQNASPSKLSKELLQSDMRVREJA	193
Qy	185	MVEQOISKLKKTQOQOLEEBAKPEBEKVPSPBESKHSRLVOIUYDENRKKAEAAHRI	244
Db	194	KYEOQLIKLKKKQOOLEEBAKPEBEKVPSPPVQKRSIVQIUYDENRKKAEBAHRI	253
Qy	245	LEGLGQVZLPLYNQSPDTRYOHENIKTIQAMRKLIIFYKRNARAKOMQYFCORYO	304
Db	254	PEGLGKQVZLPLYNQSPDTRYHENTKIQVMRKLIIFYKRNARAKOREKTCORYO	313
Qy	305	LMEALEKVEKRIENNRRAKASKVAYEYKOFPEIRKORELQOERMQSVGQSGLSMS	364
Db	314	LMEAEKQVZLRIENNRRAKASKSRIREYKQFPEIRKOREQOERNQ-RVQORGAGLSMT	372
Qy	365	AARSEHVESEIIDGSEQNLKORQOLAVIPMLYDQORIKFITNNGLMADPMKYK	424
Db	373	IARSEHSEIIEIDGSEQNNKQMLSVIPEMMDAQORVKFTINNGLMEDPMKYK	432
Qy	425	DQVNMNMSEQKEKEREKFMQHPKMGFLIASELEKTVAEACVLYYLTKKNENKSLVR	484
Db	433	DQOPFNMVMDHKEIFKDKFIQHPKNFGILASTLERKSPDCVLYLTKKNENKALVR	492
Qy	485	RSY-RRGKSQOQOQOQOQOQOQOQPMRSSHQEBKEKEKEKAEK--EERKEVEND	541
Db	493	RNYGRRGRGNQOIAAPSOEKEVEK-----EEDKAEKTEKEEKKODEKDKEDS	544
Qy	542	KEDLKEKTDUSGSDNDEKEAVASKGRKTANSQGRKRIITRSMANE---ANSEALTP	598
Db	545	KETTEKORTKTAABEPREBOVTPRGKRTANSQGGKRVTRSMTEAAMAAAATE	604
Qy	539	QOQAEI-----ASMELNRSRWTEEMETAKGGLLEHGRNMSALARMGSKTVQCKNF	652
Db	605	EPPEPLPPEPEPSTIEPVETSRWTEEMEVAKGLVHGRRMAALIAKVGTISEAQCKNF	664
Qy	653	YENYKRRONLDEIIOOHLKMEKERNARRKKKKAAPAAASEAAPPVVEDEMEASGVSG	712
Db	665	YFNYYRRRHLDMLDOOHQOKASRKRRREEDVSQCSVASYSA-----ODEDEIAS---	716
Qy	713	NEEMVEEAELHASGENVEPRGCGSPATVNNSSPTESIPSPH--TEAAK---DTQONGPK	768
Db	717	NBEENPEDESEG-----AENSSDTEASPSISPVAAKSSSDSSSENA--	756
Qy	769	PPATIGADGPPPPGPTPPRTSRABIEPTPAEATGAPTPPAPSPSAPPPVVEKEKE	828
Db	757	-----ASRGNTPEVALELEATTDAPCASP--SSAVPTTKPAERES	794
Qy	829	EE-----TAAAPVEEGEQ-----KPAAEELAVDTKABEVPVSECTEAB	871
Db	795	VEAQVTDASAEIAPEMVDHDEGCAEGSSVULDPAPYTK--ADSVDEPMQV-PENTYASKG	851
Qy	872	EGPAKQDAEAAEATBEGALNAEKKEGSGGATTAKSSGAPQ--DSDSSATCSADEVDE	928
Db	852	EGDAKERDLEST-----SEKTEARDEDDVVAQIEREPQGDSDSSATCSADEGVD	902
Qy	929	AEGGQKXNL--LSPPSLITPTGDRANASQKP--LDLKQIKQAAAIPI-----	976
Db	903	GE-PEKQKVPMDAPESLITPPGSLTI--SSPKVPLDLPLQIQAALAIYIPVWISCTPCNI	960
Qy	977	-----QVTVAHEPPREDAAPTKAPAPAPPONLQPESDAPQOQSGSPGK	1022
Db	961	PIGTVSGVALYXORITAMHSALLE-----EQQRQEQVULDEGRSSISPCST	1008
Qy	1023	SRSAPAPPDKAFAAAEQKLEGGDEPCWTSGLPFVVPPEVITKASPHADPSAFYAPPGH	1082
Db	1009	SKSP--NRE-----W-----EVQOAPVH--QVITNLPBEGV	1034

Qy	1083	PLPLGHDPARVLRPRPPIISNPPELLISSAKRPISLEROIGAIISG-----MSVOLAHVY	113
Db	1035	RLP-----TRTRPRRPPPLIPSSKTTVASER--PSPF--MGGISGCTGDTYLISSNQAVP	1086
Qy	1138	SEHAKAPVPMVMGPIPLPMDPKKLAPFGVGVKQEQUSPRGOAPRPESLQVPTQOASVLRG	1197
Db	1087	QAPRPVSIVSILGLPROQESTKAPPLTYIKOEERSPRSQNSQRPGLLV-RQHCHGVRG	1144
Qy	1198	TALGSVPGSITKGIPISTRVPSDASITTRGSTITGTPA-----DVLVKGTTTRI--IG	1246
Db	1146	TL-GAVOESITIRGTPASKISVETISSLRGSITQCTPRLPOAGIPTKALVGMPSVRMIE	1204
Qy	1249	EDSPRLDGRDSDLPKGVHIVEGKGVLVSEGMASTQCSKEDGRASSGCPHEHTAPK	1308
Db	1205	ESSPERKV---REBAASKGHVIEGSGHILSDINAKN-----REGTSPRTAHMSL-K	1255
Qy	1309	RTYDMMEGRVGRVAIS-----SASIEGLMGRAPIPERHSPH-HLKEOHNGISITQGIPIR	1361
Db	1256	RSYEAVEGSIKQGMRESRPSVAPLEGLICRALP--RSPSHSDIKERTVLGSIIMQCTPR	1313
Qy	1366	SYVEAOEVLREBALIKREGPRRPPRSRLDTEAKTQALGPIKLKPAHNGVATVTEA	1422
Db	1314	APASFEDEL-KYRQIKRES-----PIKPAEGAI-----TKGR-YDG-ITTKEM	1358
Qy	1422	GSIHEIPREEL-----NHTELPLAPRLKSGSITQCTPLKYDTGASTTSKKNVDYSL	1476
Db	1359	GSIHEIPRODILIGESKTEBEVQSTRIIEGSIISQCTPIKFDNN-SGQAIKHNVSL	1417
Qy	1477	IGSPORTPRPNRLDVMD-ARALERACE-----ESLKSPPGTAASSGGSIGARPAVIV	1530
Db	1418	ITGPRL-PRGMLEIPEIKIKVERGKYEVDVKAQEPRAHNTSVSSGSPVLST--L	1472
Qy	1531	PELGRPROSPPLTYEDHGA-----PRAGLPRGSPVPMKEPRPRLOEGSLSSK-ASQORK	1584
Db	1473	HEAPRAQUSPGLYDDSSARKTRPVQNTISGSPMNR-----TSVSSSKASHERK	1525
Qy	1585	LTSIPRE---IANSPHSTVEHHHPRIISPYENLLRGVGVLYKSHIPLAFDPTSIPRG	1640
Db	1526	STLPTQRESIPAKSPVGVQDPIVSH--SPDPHHRSSAADEVYRSHLPTHLDE-AMPFH	1582
Qy	1641	IPLDAAAYULPRNHAPRYPHILYRPLINGYDUALEN-ROTIINDYTSSOONHNT	1699
Db	1583	RALDPAAAYLLOKOLSPTRPGYPSQOLY-----AMENTKOTILINDYTSSOOMOVL	1633
Qy	1700	ATAMAQRADMLRGLSPRESSLALNYAAGPRGIIIDLCVPHILYVUPRTGTPATAMDRLA	1759
Db	1634	-----REDVTRGLSPREQRLGLPYPA-TRGIIDLTMNP-TLIVPHACGISTPMDRIT	1685
Qy	1760	YLTPRAPQFSSR-HSSPLSPGCGTHLTKPTTSSSERERDRDREDRERREKSIILTST	1818
Db	1666	YIPCTQVTFPPRPNAASLISGHTHL--AAAASAREREREREKERERERERERER	1742
Qy	1819	TTVEHAP---IMRPGTEOSSGSSGSSGSSGSSSRPASHSHAHONSISPTQDA-DOOR	1874
Db	1743	ERIAAARADLIRGSGSEQ-----PDRPSHGIVRSPR-SVRTQCTIILQOR	1787
Qy	1875	PSVLHNTMGKIIITAVBSKPTVLIRSTSPV---RPAATPRATHCPGLGTLGVVP	1930
Db	1788	PSVQGTNGTISVILRPDLPTOLRIIMPLPSGGPISISQILPASRYVTADA-LAALVUDAAAS	1846
Qy	1931	TLMEPVLLPKAPR-----VAPRPRADTGAFLAKPRABAGBLEPASPSPK	1977
Db	1847	APQMDVSKTKSEKHEARLEENLSRBAVSEGOOLEOKULBEVKSVOQCVTSSALPSC	1906
Qy	1978	GSEBRPLV-----PVSGHATIAATPRAN-LADHNASPDRAPAPASDPHRE	2024
Db	1907	KAQPHASVUVEAOKDQKPRPKSYEBELRTRGKTTITTAANFIDIVITTRQIADKDARER	1966
Qy	2025	KTQSKPISIQEILRSIGYHGSIVSPGCVPRVSVSSPLTHQKGLPKHLEIDKSLLEG	2084
Db	1967	GSQSSDS-----SSLSSHYETJASALIVISPASSPAPQOEKPAOVYORDMYKAOAEN	2020

ALIGNMENTS

RESULT 1

S60255

transcription co-repressor SMRT - human

C/Species: Homo sapiens (man)

C/Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 05-Nov-1999

C/Accession: S60255

R/Chen, J.D.; Evans, R.M.

A/Title: A transcriptional co-repressor that interacts with nuclear hormone receptors.

A/Reference number: S60255; MUID:96008552; PMID:7566127

A/Accession: S60255

A/Status: preliminary; nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-1495 <CHE>

A/Cross-references: EMBL:U57146; NID:g1045654; PIDN:AAC50236.1; PID:g1045655

Query Match

Best Local Similarity 59.6%; Score 7871; DB 2; Length 1495;

Matches 1487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 729 PATAMRLAAYLPTAPQPFSSRHSSPLSPGPTHLTKPTTSSSERERDRDRDR 788

QY 1811 EKSLITSTTVERAPITWRPCTEASSGSSGSSGSSGSSSRPASHSHAHQSPISPRQDA 1870

Db 789 EKSLITSTTVERAPITWRPCTEASSGSSGSSGSSGSSSRPASHSHAHQSPISPRQDA 848

QY 1871 LQORPSVYHNTGKGGITTAVERPKPTLARTSTSSPPVPAATPPATPCPLGGLDGVVP 1930

Db 849 LQORPSVYHNTGKGGITTAVERPKPTLARTSTSSPPVPAATPPATPCPLGGLDGVVP 908

QY 1931 TIMEPVLLPKKARVAVRPERPRADTGHAFLAKPPARSGLEPASSPSKSGSEPRPLVPVSG 1990

Db 909 TIMEPVLLPKKARVAVRPERPRADTGHAFLAKPPARSGLEPASSPSKSGSEPRPLVPVSG 968

QY 1991 HATITARTPAKNLA PHNASPPPPAPASAPPHREKTKOSKFPSTIOELELRSLGTHGSSYSP 2050

Db 969 HATITARTPAKNLA PHNASPPPPAPASAPPHREKTKOSKFPSTIOELELRSLGTHGSSYSP 1028

QY 2051 EGYEPPVSPVSPSLTHDKGLPKLLEELDKSHLGEELRPKOPGVKLGGEAAHLPHLRPLP 2110

Db 1029 EGYEPPVSPVSPSLTHDKGLPKLLEELDKSHLGEELRPKOPGVKLGGEAAHLPHLRPLP 1088

QY 2111 ESQPSSEPLIQTAPGVKQHVYTLAQHISEVITODYTRHHPOQLSAPLPALYSFPGAS 2170

Db 1089 ESQPSSEPLIQTAPGVKQHVYTLAQHISEVITODYTRHHPOQLSAPLPALYSFPGAS 1148

QY 2171 CPVLDLRRPPSDLYLPPPDGAPARSGPHSEGGKRSSEPKTSTVLGGEGEIEVSPPEG 2220

Db 1149 CPVLDLRRPPSDLYLPPPDGAPARSGPHSEGGKRSSEPKTSTVLGGEGEIEVSPPEG 1208

QY 2231 MTEPGHRSAYVPLLYNDGQTEPSRMGSKSPGNTSOPPAFFSKLTESNAMYKSKQOEI 2290

Db 1209 MTEPGHRSAYVPLLYNDGQTEPSRMGSKSPGNTSOPPAFFSKLTESNAMYKSKQOEI 1268

QY 2291 NKCLNTNREPEYNISSQPTETI FNNPATITGTGLMTYRQAQVQEHASTNGLXAIIRKAL 2350

Db 1269 NKCLNTNREPEYNISSQPTETI FNNPATITGTGLMTYRQAQVQEHASTNGLXAIIRKAL 1328

QY 2351 MGKYDQWEEBPLSANAFNPLNLSASLPAAMPITTAADGRDHTLTSQGGGKXAVSRRPS 2410

Db 1329 MGKYDQWEEBPLSANAFNPLNLSASLPAAMPITTAADGRDHTLTSQGGGKXAVSRRPS 1388

QY 2411 SRRKSPAPGLAGDRPPSVSVHSEDCNRRPPLTNRWEDRPSAGSTPPFPNPLMR 2470

Db 1389 SRRKSPAPGLAGDRPPSVSVHSEDCNRRPPLTNRWEDRPSAGSTPPFPNPLMR 1448

QY 2471 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1449 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

Db 1495 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 1495

QY 2517 LQGVNASPPPPGLPAGSGPLAGPHHAMDEEPKPLCSQYETLSDSE 2517

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2004, 21:53:54 ; Search time 60 Seconds
(without alignments)
4035.235 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 1215
Sequence: 1 MSGSTQLVAQTRATEPRYP.....WDEPKPLCSQYETLDSSE 2517

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7871	59.6	1495	2	transcription co-r
2	4187	31.7	2453	2	nuclear receptor c
3	1272	9.6	1047	2	hypothetical prote
4	574	4.3	5327	2	microtubule-associ
5	552	4.2	2649	2	hypothetical prote
6	526	4.0	2142	2	MHC class III hist
7	520	3.9	2187	2	nascent polypeptid
8	512	3.9	5762	2	proline-rich pepti
9	508.5	3.8	1872	2	MHC class III hist
10	504	3.8	3942	2	Bassoon protein -
11	503.5	3.8	7962	2	elastic titin - hu
12	503	3.8	1870	2	MHC class III hist
13	502.5	3.8	2715	2	eyelid - fruit fly
14	498	3.8	3938	2	Bassoon protein -
15	492.5	3.7	2774	2	microtubule-associ
16	477	3.6	2282	2	DNA-binding protei
17	470.5	3.6	4957	2	AKR protein - huma
18	470.5	3.6	5262	2	hypothetical prote
19	445.5	3.4	1952	2	collagen alpha 1(I
20	436	3.3	1151	2	microtubule-associ
21	424.5	3.2	2464	1	high molecular mas
22	419	3.2	5170	2	microtubule-associ
23	416.5	3.2	1791	2	hypothetical prote
24	412.5	3.1	1226	2	hypothetical prote
25	407.5	3.1	1464	1	collagen alpha 1(I
26	406.5	3.1	2364	2	microtubule-associ
27	403	3.1	1274	2	hypothetical prote
28	404	3.0	1110	2	NF-160 - sea lampy
29	397	3.0	1670	1	collagen alpha 3(I

30	395	3.0	1520	2	T00273	hypothetical prote
31	394	3.0	2783	1	A41948	alpha-fetoprotein
32	391	3.0	3924	2	S37431	ankyrin 2, neuron
33	389.5	2.9	1188	2	S49915	extensin-like prot
34	389.5	2.9	3869	2	A48205	All-1 protein +GFE
35	389	2.9	1320	2	JC5630	TCOF1 protein - mo
36	383.5	2.9	3968	2	A44265	trithorax homolog
37	383	2.9	2944	2	A54849	collagen alpha 1(V
38	377	2.9	1691	1	S22917	collagen alpha 5(I
39	375	2.8	1006	2	T42721	atrophin-1 related
40	374.5	2.8	2722	2	T20532	hypothetical prote
41	373.5	2.8	1184	2	S50832	atrophin-1 - huma
42	373.5	2.8	1494	2	T14355	protein-tyrosine-p
43	373.5	2.8	2938	2	T30249	cell proliferation
44	373	2.8	1763	2	S16366	collagen alpha 2(I
45	372.5	2.8	1173	2	T31421	C-terminal domain-
46	372.5	2.8	1184	2	G01763	atrophin-1 - huma
47	368.5	2.8	2738	2	E88320	protein F07A11.6 I
48	367	2.8	1611	2	T38236	hypothetical prote
49	360	2.7	2232	2	T34434	hypothetical prote
50	358	2.7	1618	2	S21424	nestin - human
51	356.5	2.7	2526	2	T20531	hypothetical prote
52	355	2.7	2090	2	S26058	probable transform
53	354	2.7	3507	2	T34513	hypothetical prote
54	350.5	2.7	2843	1	RBHUP	adenomatous poly
55	349.5	2.6	1840	2	T30250	GM1 protein - mous
56	348	2.6	1669	1	CGMS4B	collagen alpha 1(I
57	348	2.6	3256	2	A48666	cell proliferation
58	346.5	2.6	1018	2	S44758	CI4B9.6 protein -
59	345	2.6	1843	2	S18803	collagen alpha 1(V
60	344.5	2.6	1546	1	CGHUE	collagen alpha 2(X
61	344	2.6	2845	2	I49505	adenomatous poly
62	343.5	2.6	1464	1	S59856	collagen alpha 1(I
63	343	2.6	1466	1	CGHUTL	collagen alpha 1(I
64	342.5	2.6	1020	1	QPHDH	neurofilament trip
65	342	2.6	5105	2	T32650	hypothetical prote
66	341	2.6	1200	2	A46194	neurofilament prot
67	341	2.6	1453	2	S21626	collagen alpha 1(I
68	341	2.6	1571	2	T00062	hypothetical prote
69	339.5	2.6	1690	1	CGHUIB	collagen alpha 4(I
70	338.5	2.6	2897	2	B48666	cell proliferation
71	338	2.6	13055	2	T16580	hypothetical prote
72	337	2.6	1492	2	A40333	collagen alpha 1(I
73	337	2.6	1634	2	T26517	hypothetical prote
74	336	2.5	839	2	F75518	hypothetical prote
75	335.5	2.5	3570	2	T45025	myosin I precursor,
76	335	2.5	1344	1	A35175	myosin I precursor,
77	335	2.5	1418	2	T45467	collagen alpha 1(I
78	335	2.5	1460	1	EDBRIF	immediate-early pr
79	333.5	2.5	1487	1	CGHUEC	collagen alpha 1(I
80	331	2.5	1838	1	CGHUIV	collagen alpha 1(V
81	330	2.5	1367	1	S48478	glucan 1,4-alpha-g
82	329.5	2.5	916	2	A27864	neurofilament trip
83	329.5	2.5	1819	2	T32008	hypothetical prote
84	329	2.5	1280	2	T00365	hypothetical prote
85	326.5	2.5	1414	1	S23809	collagen alpha 2(I
86	326.5	2.5	1469	2	T09219	basal transcrip
87	325	2.5	1829	2	T26135	hypothetical prote
88	324.5	2.5	2022	2	T48818	hypothetical prote
89	324	2.5	3149	1	QOBER	glucan 1,4-alpha-g
90	323.5	2.4	1497	2	I49607	BPLF1 protein - hu
91	323.5	2.4	1603	2	S23810	procollagen type V
92	323.5	2.4	4549	2	T20771	collagen alpha 1(X
93	323.5	2.4	4667	2	T20774	hypothetical prote
94	323	2.4	1446	1	A45344	hypothetical prote
95	322.5	2.4	1684	2	JW0057	immediate-early pr
96	320.5	2.4	1669	1	CGHUEB	collagen alpha 1(I
97	319	2.4	2361	2	T25752	hypothetical prote
98	318.5	2.4	1094	2	S49313	protein kinase - s
99	318	2.4	1924	2	S27923	gene lrf protein -
100	318	2.4	1467	2	B41182	collagen alpha 1(I